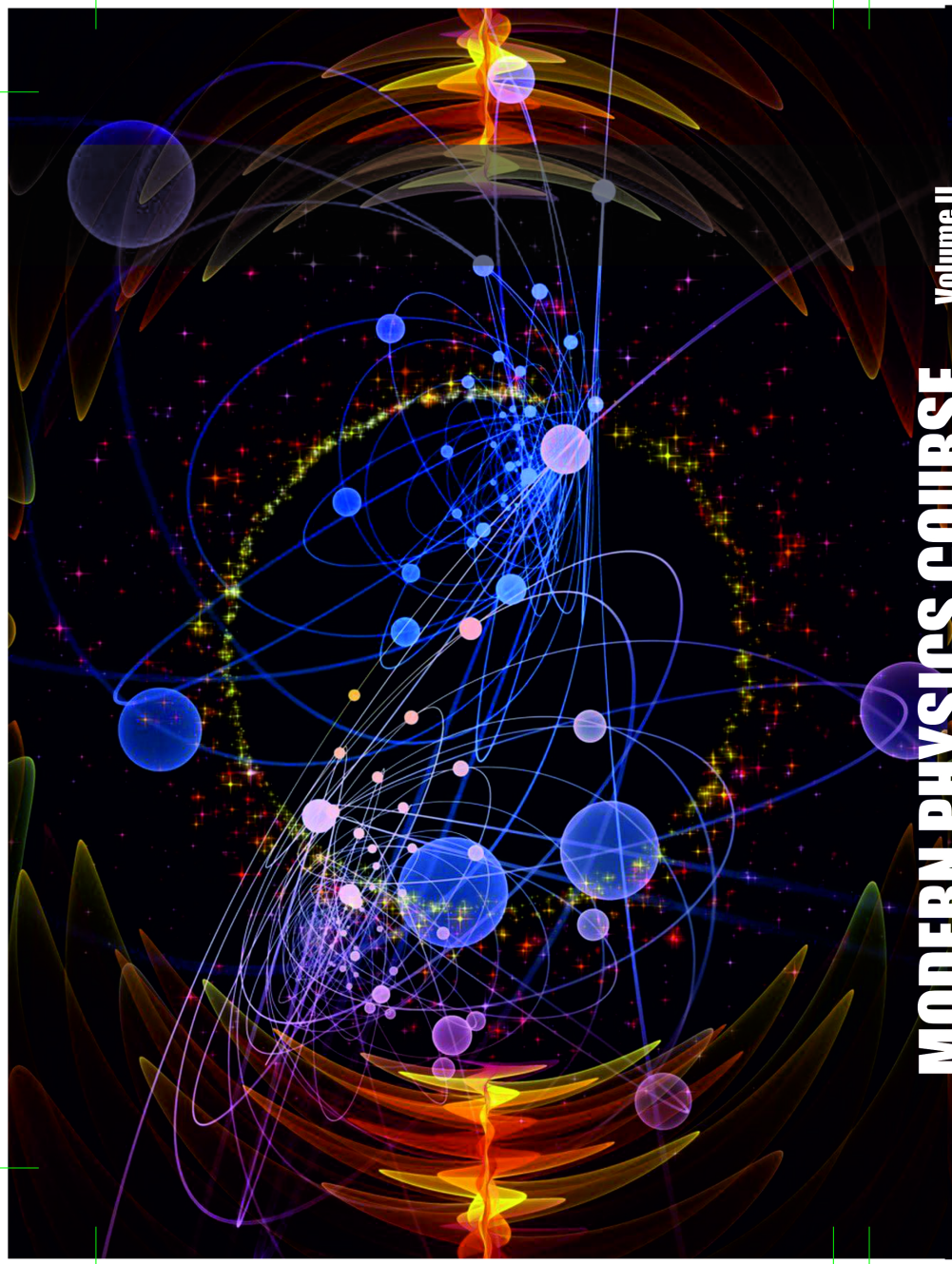


MODERN PHYSICS COURSE

Volume II

L. Preigerman, M. Brooke, V. Mikhalev-Volookov

MODERN PHYSICS COURSE

New Approachesto Explaining
the Physical Picture of the World

Universe

*

Movement

*

Force fields

*

Oscillations and waves

*

Fundamentals of quantum physics

Volume II

Israeli Independent Academy for the Advancement of Science
L. Preigerman
M. Brooke

MODERN PHYSICS COURSE

Under the editorship of
Doctor of Philosophy Professor O.E. Baksanskiy

in 5 parts
Third Edition

Part 4. Oscillations and waves

Recommended by the Academic Council of the Russian State Classical Academy named after Maimonides (Rambam) for undergraduate and graduate students, and for secondary and high school physics teachers.

Volume 2

Moscow URSS 2014

Table of Contents

Chapter 5. Oscillatory processes

Oscillatory processes or simply **oscillations** are such forms of motion that have a repeatability in time. Oscillations are called **periodic** if the physical quantities characterizing them are repeated at equal intervals of time. All oscillations, regardless of their physical nature, are described by the same equations, which differ from each other only by their constituent physical quantities that characterize the observed oscillatory process.

5.1. Types of oscillations

A distinction is made between mechanical and non-mechanical oscillations. The **mechanical** ones, in particular, include the above-mentioned oscillations of systems in the field of action of elastic and gravitational forces. These are, for example, oscillations of a mathematical or physical pendulum under the action of gravity, oscillations of a spring under the action of elastic forces, etc. Mechanical oscillations also include periodic movements of bodies along closed paths (circle, ellipse, etc.) under the action of central forces, piston motion in an internal combustion engine, vibrations of strings, air and gases, various kinds of vibrations, etc. The sources of mechanical oscillations are, as a rule, the oscillations of bodies (vibrators, membranes, etc.), continuous media (liquids, gases) as well as structural elements (particles) of physical systems. The latter include, for example, the oscillatory motion of gas molecules, free electrons or other electric charges in conductors of various kinds, etc.

Non-mechanical oscillations refer to the physical characteristics of more complex forms of motion. Their sources are repetitive changes in time of both mechanical and non-mechanical movements.

For example, the sources of electromagnetic oscillations are periodically changing in time mechanical movements of electric charges in conductors or changes of electric and magnetic values of the corresponding fields. The sources of climatic and other types of

oscillations of atmospheric and lithospheric processes are periodically changing values of temperature, pressure, etc. The sources of oscillations of vital activity of organisms are rhythmically repeated biological processes, and the sources of oscillations of radiation and radioactive background and cyclically changing intra-atomic or intranuclear processes are periodically repeated interactions, as well as the birth and annihilation of elementary particles, etc.

Depending on the nature of the impact on the system of external factors oscillations are divided into natural (free), forced, parametric and self-renewing (self-oscillation).

The **natural, or normal mode** oscillations are the oscillations that occur in the system, in the absence of influence on it of external factors. The normal mode oscillations are, for example, those of various kinds of pendulums left to themselves, as well as the oscillations of electric and magnetic fields in electric circuits in the absence of external non-electric sources.

Forced oscillations refer to oscillations that occur under the influence of periodically changing external factors. For example, the oscillation of a swing under the action of rocking impulses produced by humans, or the oscillation of electric and magnetic fields in AC electric circuits initiated by sources of external non-electrical energy.

Self-oscillation is an oscillating process which, like forced oscillations, occurs under the action of external energy introduced into the system. The nature of this energy input differs,

however, in that its source is commutated by feedback and, by compensating for the energy of dissipation synchronously with the oscillations of the system, keeps the oscillations in it unchanged. Self-oscillations occur, for example, in mechanical clocks, electromagnetic oscillators, as well as in the heart muscle.

Parametric oscillations are such forced oscillations, under the action of which there is a periodic change of any system parameter. For example, a periodic change in the length of a mathematical pendulum, or the capacitance (inductance) in an alternating current circuit, etc.

Systems in which oscillatory processes are possible are divided into **one-dimensional systems (oscillators) and multidimensional systems**. One-dimensional systems are those having one degree of freedom and characterized, respectively, by one generalized coordinate q. Systems with two degrees of freedom can be distinguished among multidimensional systems. The oscillations of a multidimensional system can be viewed as the result of addition of several oscillations of one-dimensional systems. These are, for example, resultant oscillations of systems consisting of one-dimensional oscillators that make oscillations in mutually perpendicular directions, etc.

Oscillations of systems are divided into **linear**, described by linear differential equations, and non-linear, described by **non-linear** equations.

The character of oscillations is uniquely described by the law of oscillatory motion.

$$q_i = q(t) \tag{5.1}$$

When the law of oscillatory motion is given by trigonometric functions $Sin\ \alpha t$ or $Cos\ \alpha t$, oscillations are called **harmonic**. In all other cases, oscillations are called **non-harmonic** or **complex**. Systems that perform harmonic oscillations are called **harmonic oscillators**.

In mathematics it is proved that any complex oscillation can be represented as a result of addition of harmonic oscillations. The specified sum of harmonic oscillations is described by the Fourier series (Appendix 2).

In the process of oscillation there is a transformation of some kinds of energy into other kinds of energy. For example, in mechanical oscillations, kinetic energy is converted into potential energy and back. During electromagnetic oscillations there is a mutual transformation of electric and magnetic energy.

If the natural oscillations occur without losses (dissipation) of energy, they are called **undamped**. According to the law of conservation of energy, undamped oscillations, once created, continue indefinitely. Undamped natural oscillations are an idealization of real oscillations,

which never occur without energy dissipation. The natural vibrations of real systems are always damped.

An **equilibrium** state is a state of a system that does not change over time. The motion of a system near an equilibrium state is the result of its small deviations (perturbations) from equilibrium. For one-dimensional systems, equilibrium is **stable** if, under a small perturbation (deviation), the system returns to it by making damped free oscillations relative to it.

Mechanical systems, as was shown above (Section 4.1.1.2), are in a state of stable equilibrium if their potential energy in this state is minimal. In the case of an **unstable equilibrium**, the system deviates from it further and further over time, and the deviations increase aperiodically. For multidimensional systems, the motion near the equilibrium state is more complex and depends to a large extent on the initial conditions. In the case when small deviations of the system at any smallness translate the system from one state of equilibrium to another, the equilibrium is called **indifferent**.

5.1.1. Natural undamped oscillat0ions

The equation of motion of a system performing oscillations, generalized to any form of motion, is called the equation of oscillation.

Let us consider the equation of natural mechanical oscillations in the absence of friction. Let us choose an inertial reference system and assume, for simplicity, that the oscillations occur with negligible amplitude in such a one-dimensional system, which can be identified with a material point. Let us place the origin of coordinates at the equilibrium position of the system. Let us direct the Ox axis along the deviation of the material point from the equilibrium position. Then the current coordinate at a given time x(t) coincides with the amount of deviation of the system. Let us assume that at the initial moment of time $t = t_0 = 0$ the system is in a state of stable equilibrium. Then, when the system deviates, there is a force arises that is directed against the deviation. In other words, in the direction of the Ox axis

$$F_x = kx, \tag{5.2}$$

where k is the transfer mobility of the system.

For small deviations, the relation (5.2) is valid for any mechanical system. The force counteracting the deflections is called quasi-elastic. Since the oscillations are free, according to the second law of dynamics

$$m\ddot{x} = -kx \tag{5.3}$$

Equation (5.3) can be generalized to any natural undamped oscillatory processes. In this case

$$L\ddot{q} = -\frac{1}{C}q, \tag{5.4}$$

where q is the generalized coordinate of one-dimensional oscillations;

L is a quantity characterizing the degree of inertness of the system, counteracting any possible changes in its state, and is equivalent to the mass of the mechanical system;

$\ddot{q}$ is the generalized rate of deviation of the system state from equilibrium, equivalent to the acceleration of the mechanical system;

$\frac{1}{C}$ is the coefficient of proportionality between the generalized force and the value of deviation of the system from equilibrium, equivalent to transfer mobility.

Let us write (5.4) in the following form

$$\ddot{q} + \frac{1}{LC}q = 0 \tag{5.5}$$

The relation (5.5) is a generalized equation of small one-dimensional free oscillations of non-dissipative systems. In mathematical terms, this is a linear one-dimensional second-order differential equation with constant coefficients (Appendix 2). Therefore, the oscillations it describes are linear. Its solution has the form

$$q = C_1 e^{r_1 t} + C_2 e^{r_2 t} \tag{5.6}$$

where C_1 and C_2 are integration constants;

r_1 and r_2 are roots of the characteristic equation, which is obtained from the differential equation by replacing the derivatives by the corresponding degrees

$$r^2 + \frac{1}{LC} r = 0 \tag{5.7}$$

Let us denote

$$\frac{1}{LC} = \omega_0^2 \tag{5.8}$$

Then

$$r^2 + \omega_0^2 r^2 = 0 \tag{5.9}$$

The roots of equation (5.9) are

$$r_{1,2} = \pm j\omega_0 \tag{5.10}$$

Substituting (5.10) into (5.6) gives

$$q = C_1 e^{j\omega_0 t} + C_2 e^{j\omega_0 t} \tag{5.11}$$

If we assume that at the initial moment of time $t = t_0 = 0$ and accordingly,

$$\left.\begin{array}{l} q = q_m \\ \dot{q} = 0 \end{array}\right\} \tag{5.12}$$

then

$$\left.\begin{array}{l} C_1 + C_2 = q_m \\ C_1 - C_2 = 0 \end{array}\right\}$$

Determining C_1 and C_2 from this system and substituting them into (5.11), we obtain

$$q = q_m \frac{e^{jw_0 t} + e^{-j\omega_0 t}}{2} \tag{5.13}$$

From (5.13) and the Euler formula (Appendix 5) it follows that

$$q = q_m \cos \omega_0 t \tag{5.14}$$

Solutions and their linear combinations of the following form also satisfy equation (5.5)

$$q = q_m \cos(\omega_0 t + \varphi_0) \text{ and } q = q_m \sin(\omega_0 t + \varphi_0)$$

*where φ_0 is a constant.

It follows from (5.14) that the value $q = q_m$ is the maximum deviation of the system from the state of equilibrium. In the theory of oscillations the value q_m is called the amplitude (see Section 2.6.1).

The value of the generalized coordinate q is expressed through the function $\cos \varphi$. This means that the one-dimensional natural oscillations at small deviations of the system are harmonic.

It is known that $\cos \varphi$ is a periodically changing function with period 2π, i.e.

$$\cos(\varphi + 2\pi n) = \cos \varphi, \text{ where n = 1, 2, 3...} \tag{5.15}$$

It is also known (see Section 2.6.1) that the time interval t = T at which the quantity q repeats its value, characterizes the period of change of the physical quantity in question.

From the equation

$$\cos \omega_0 T = 1 \tag{5.16}$$

it follows that

$$\omega_0 = \frac{2\pi}{T} = 2\pi v \tag{5.17}$$

where ω_0 is the cyclic frequency of oscillations;

v is the linear frequency of oscillations.

$$v = \frac{1}{T} \tag{5.18}$$

For free oscillations the value of o0 is called the **natural frequency** of oscillations.

If T = 1, then

$$v = 1.$$

This unit (see Section 2.6.1) is called Hertz (Hz). From this definition, it follows that 1 Hz is the frequency of oscillation equal to one oscillation per second

$$1Hz = 1\frac{osc.}{second} \tag{5.19}$$

It follows from (5.17) and (5.18) that

$$\omega_0 = 2\pi v \tag{5.20}$$

On the other hand, according to (5.14)

$$\omega_0 = \frac{\varphi}{t} \tag{5.21}$$

where φ is the argument value of the harmonic function;

The value ω_0 determines the rate of change of the angular argument φ over time. It is therefore also called the **angular frequency**. For natural oscillations, the angular frequency is called the **natural** frequency. It also follows from (5.21) that the value

$$\varphi = 2\pi\frac{t}{T} \tag{5.22}$$

It is clear from (5.22) that the quantity φ, called the oscillation **phase**, determines a part of the oscillation period of the system elapsed up to the time t (see also Section 2.6.1). For example, if

$$\varphi = n\frac{\pi}{2}$$

then

$$n = \frac{\pi}{2} = 2\pi n\frac{t}{T};\ t = \frac{T}{4}$$

This means that since the beginning of the next oscillation period, the system has passed its fourth part. The phase of the oscillations at the time $t = t_0$ is called the **initial phase**, so in the general case

$$q = q_m \cos(\omega_0 t + \varphi_0) \tag{5.23}$$

where $q_m = A_0$ is the amplitude of oscillations;

φ_0 is the initial phase.

It follows from (5.23) that the speed of one-dimensional harmonic oscillations

$$q = v_m \cos(\omega_0 t + \varphi_0) \tag{5.24}$$

where v_m is the amplitude value of the velocity.

$$v_m = -\omega_0 q_m$$

The minus sign means that the velocity is always directed against the deflection of q (if q increases, the velocity decreases, and vice versa).

Since at $t = t^0 = 0$

$$q_0 = q_m \cos\varphi_0;\ \dot{q}_m = -\omega_0 q_m \sin\varphi_0$$

then

$$\varphi_0 = -\tan^{-1}\frac{\dot{q}_0}{\omega_0 q_0};\ q_m = q_0\sqrt{1 + \frac{\dot{q}_0^2}{q_0^2 + \omega_0^2}}$$

From field theory (Appendix 3) it is known that for potential fields the generalized force Q

$$Q = -\frac{dU}{dq} \tag{5.25}$$

where U is the interaction energy of the field sources and test charges.

Since, according to (5.4)

$$Q = -\frac{1}{c}q$$

Then

$$U = \int \frac{1}{C} q dq \tag{5.26}$$

By integrating (4.20), we obtain that

$$U = \frac{q^2}{2C} + B \tag{5.27}$$

If we assume that $q = 0$ when $U = 0$, $V = 0$, then

$$U = \frac{q^2}{2\mathrm{C}} \tag{5.28}$$

Total energy of the mechanical system

$$\varepsilon = \varepsilon_k + U; \varepsilon = \frac{L\dot{q}^2}{2} + \frac{q^2}{2C} \tag{5.29}$$

Substituting q *and* $\dot{q}$ from (5.23) and (5.24) in (5.29) gives

$$\varepsilon = \frac{LC\omega_0^w q_m^2 \sin^2(\omega_0 t + \varphi_0) + q_m^2 \cos^2(\omega_0 t + \varphi_0)}{2C} \tag{5.30}$$

Since by definition $\omega_0^2 = 1/$ LC, the

$$\varepsilon = \frac{q_m^2[\sin^2(\omega_0 t + \varphi_0) + \cos^2(\omega_0 t + \varphi_0)]}{2C} \tag{5.31}$$

Total energy of small oscillations in a potential field

$$\varepsilon = \frac{q_m^2}{2C} \tag{5.32}$$

In other words, the total energy of an oscillating non-dissipative conservative system is proportional to the square of the amplitude of the system's deviation q_m from the state of equilibrium.

5.1.2. Natural damped oscillations

In the case of mechanical oscillations, for which energy dissipation is caused by friction, we can assume that the frictional force is proportional to speed

$$F_{TP} = -f\dot{x} \tag{5.33}$$

where *f* is the generalized coefficient of friction.

The equation of motion in the absence of external forces takes the form

$$m\ddot{x}+fx+kx=0 \tag{5.34}$$

or

$$\ddot{x} + \frac{f}{m}\dot{x} + \frac{k}{m}x = 0 \tag{5.35}$$

For any form of motion, the equation of oscillations of the type in question can be obtained by generalizing equation (5.35)

$$\ddot{q} + \frac{R}{L}\dot{q} + \frac{1}{LC}q = 0 \tag{5.36}$$

where *R* is the value of the generalized resistance, characterizing the degree of dissipation of the system, equivalent to the friction coefficient *f*. The quantity *R* is often called the active or dissipative resistance of the system. If we denote

$$\frac{R}{L} = 2\delta \tag{5.37}$$

$$\frac{1}{LC} = \omega_0^2 \tag{5.38}$$

then the equation of oscillation takes the form

$$\ddot{q} + 2\delta\dot{q} + \omega_0^2 q = 0 \tag{5.39}$$

This equation, as stated above, has the following solution

$$q = C_1 e^{r_1 t} + C_2 e^{r_2 t} \tag{5.40}$$

where $r_{1,2}$ are roots of the characteristic equation

$$r^2 + 2\delta r + \omega_0^2 = 0 \tag{5.41}$$

and its solutions

$$r_{1,2} = -\delta \pm \sqrt{\delta^2 - \omega_0^2} \tag{5.42}$$

C_1 and C_2 are integration constants, which are found taking into account the initial conditions.

Let us denote the quantities $\omega L = X_L$ и $\frac{1}{\omega C} = X_C$. These quantities have the dimension and meaning of resistance. However, unlike the active resistance, which characterizes the degree of energy dissipation, the X_L and X_C resistances determine the phase of oscillation and the degree of its shift. That is why they are called **reactive**. If we assume that when $t = t_0 = 0$, $q = q_0$, and $\dot{q}_0 = 0$ when $\delta < \omega_0$ ($R < 2\sqrt{X_L X_C}$), then

$$q = A_0 e^{-\delta t} \cos(\omega_0 t + \varphi_0) \tag{5.43}$$

where A_0 is the initial amplitude of oscillations;

φ_0 is the initial phase of the oscillations;

ω is the natural frequency of the dissipative system.

At the same time

$$A_0 = \frac{q_0}{\sqrt{1 - \frac{R^2 C}{4L}}} \tag{5.44}$$

$$\varphi_0 = \tan^{-1} \sqrt{\frac{4L}{R^2 C} - 1} \tag{5.45}$$

$$\delta = \frac{R}{2L} \tag{5.46}$$

$$\omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}, \left(\omega = \sqrt{\omega_0^2 - \delta^2}\right) \tag{5.47}$$

In the case when $\delta > \omega_0$ ($R > \omega_0\sqrt{X_L X_C}$, the oscillatory process of the dissipative system degenerates into an aperiodic process (see Appendix 2) and

$$q = A_0 e^{-\delta t} sh(\omega t + \varphi_0) \tag{5.48}$$

It follows directly from equations (5.43) and (5.46) that the quantity

$$A = A_0 e^{-\delta t} \tag{5.49}$$

is the amplitude of the oscillations. Thus, the attenuation of the amplitude is higher the greater the dissipation and the smaller the inertia of the system (the value L). If the active resistance is very large compared to the reactive resistance, the process becomes aperiodic. In an oscillatory damping process, the amplitude decreases exponentially $e^{-\delta t}$ (Fig. 5.2)

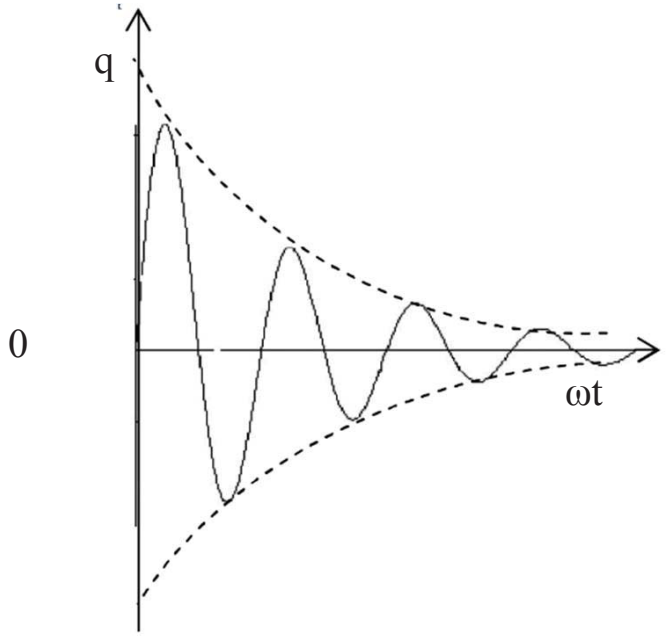


Figure 5.2.

The dotted lines in the figure are envelopes and represent the exponential law of attenuation of the amplitude of oscillations.

The damping intensity of an oscillatory process is defined as the amount of attenuation of the amplitude in one period. This quantity is usually characterized by the ratio of the amplitude A(t) at a given time to the amplitude A(t + T) through the period of oscillation T. However, in order to increase the clarity of the specified characteristic of damping, we do not consider the ratio itself, but its natural logarithm.

The attenuation characteristic, therefore, is called the **logarithmic decrement of damping** χ.

It is obvious that

$$\chi = \ln\frac{A(t)}{A(t+T)} = \ln\frac{e^{-\delta t}}{e^{-\delta(t+T)}}\delta T \tag{5.50}$$

From (5.50) we see that the logarithmic decrement of damping is higher the greater the dissipation and period of oscillation.

Another characteristic of the intensity of the damping of oscillations of the system is the relaxation time, equal to the time interval τ , during which the amplitude of oscillations decreases by *e* times. Since

$$\frac{e^{-\delta t}}{e^{-\delta(t+\pi)}} = e \tag{5.51}$$

Then

$$\delta\tau = 1;\ \tau = \frac{1}{\delta} \tag{5.52}$$

It is, in this way,

$$A(t) = A_0 e^{\frac{t}{\tau}} \tag{5.53}$$

As an example of a system performing non-mechanical natural damped oscillations, let us consider an **oscillating circuit**. It represents an electric circuit (Fig. 5.3), consisting of a series-connected solenoid 1,

in the form of an inductance coil L, and a capacitor 2 with capacity C. Each of these elements has an active resistance, on which energy dissipation occurs. To simplify things, let's assume that the active resistances of the circuit elements are zero, and their actual active resistance R is concentrated in a separate element, resistor 3, included in series with the other elements in the circuit. Let us also assume that the circuit capacitor is charged from the source of e.m.f. 4 to the value q and at the initial moment $t = t_0$ it is switched off by switch 5.

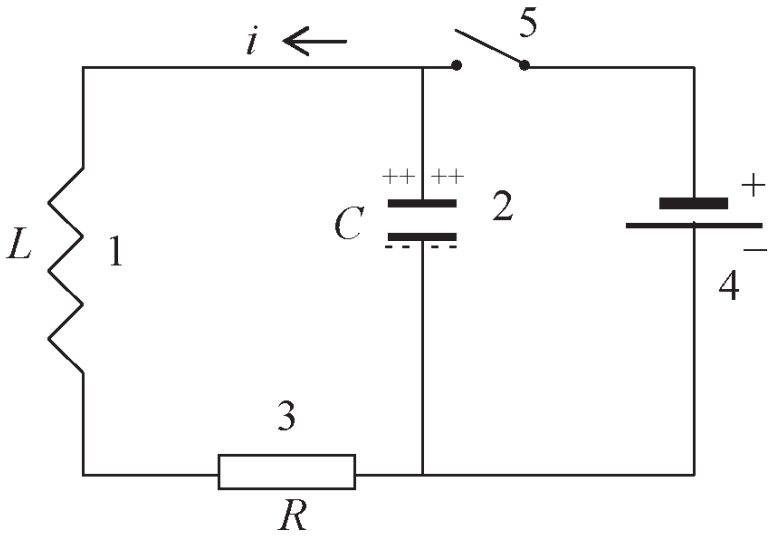


Figure 5.

At the initial moment, the energy of the circuit is completely concentrated in the capacitor 2 and is the energy of the electric field intensity $\vec{E}$ between its plates. After disconnecting the source of emf 4 negative charges under the action of the electric field (potential difference) move from the bottom plate of the capacitor to the top plate, creating a current i in the circuit. A magnetic field of intensity $\vec{H}$ arises around conductor 1 with inductance L. This field creates a magnetic flux Φ in the inductance that is proportional to the current i

$$\Phi = Li \tag{5.54}$$

As charges move from one capacitor plate to another, the capacitor is recharged and the current, having passed through the maximum, becomes zero. The magnetic flux passes along the axis of the

inductor coil and is concentrated inside it. The magnetic flux in the connecting wires due to its smallness can be neglected. Since the current i is continuously changing, then according to (5.54) the magnetic flux changes accordingly

$$\frac{d\Phi}{dt} = L\frac{di}{dt} \tag{5.55}$$

Variable magnetic flux in the coil, in turn, leads to the appearance of self-induction current i_1, which, according to Lenz's law, is directed against the main current i, and counteracts its increase. As the current i increases, the capacitor discharges, and the maximum electric field strength $\vec{E}$ in the initial state decreases to zero. At the same time, the magnetic field $\vec{H}$ in the inductor coil grows, and when the electric field strength decreases to zero, $\vec{E} = 0$, the strength $\vec{H} = \vec{H}_{max}$. This means that the energy of the electric field is completely converted into the energy of the magnetic field. The current i, having reached the maximum value, begins to decrease, recharging the capacitor. An electric field of the opposite direction is established between the plates of the capacitor. As a result, the current and magnetic field strength decreases to zero, and the electric field strength increases to a negative maximum. As a result of the conversion of magnetic energy into electrical energy, the process reverses its direction and goes to the initial state. Then it is repeated all over again. Thus, the source of the oscillatory process in the oscillating circuit is the oscillatory motion of electrons in the conductors of the electric circuit, which excite periodic changes in the electric and magnetic field strengths of the circuit and lead to the appearance of **electromagnetic oscillations** (see section 4.2.3). In the case when

$$\left.\begin{array}{l} R \ll X_L \\ R \ll X_C \end{array}\right\} \tag{5.56}$$

dissipation of electromagnetic energy is insignificant and electromagnetic oscillations in the circuit can continue long enough, gradually damping. If ideally we assume that $R \approx 0$ then the oscillations obey equation (5.5), i.e. they are harmonic and undamped.

Electromagnetic oscillations in a real oscillating circuit (R ≠ 0) occur with dissipation of electromagnetic energy due to its conversion into thermal energy and are natural damped oscillations. They obey equation (5.36) and its solution (5.40) taking into account (5.37) and (5.38), provided that

$$R < 2\sqrt{X_L X_C} \tag{5.57}$$

5.1.3. Forced oscillations

The equation of forced oscillations can be obtained by adding to the homogeneous differential equation (5.34) or (5.36) of natural oscillations the right part in the form of a generalized periodic external force F(t). Then the equation takes the form

$$\ddot{q} + \frac{R}{L}\dot{q} + \frac{1}{LC}q = \frac{1}{L}F(t) \tag{5.58}$$

Equation (5.58) is a linear inhomogeneous differential equation of the second order with constant coefficients. According to the theory of differential equations (Appendix 2), the solution of equation (5.58) has the form

$$q = q_1 + q_2, \tag{5.59}$$

where q_1 is the general solution of the homogeneous equation according to (5.43);

q_2 is a partial solution of an inhomogeneous equation.

Let F(t) be a harmonic function, i.e.

$$F(t) = F_0 \cos \Omega t, \tag{5.60}$$

where Ω is the oscillation frequency of the external driving force.

As a result, the solution of equation (5.58) for the steady state takes the form

$$q = q_2 = A\cos(\Omega t + \varphi_1),$$

where A and φ_1 are arbitrary constants.

Substituting this solution into (5.58) gives

$$A = \frac{F_0}{L\sqrt{(\omega_0^2 - \Omega^2)^2 + 4\delta^2\Omega^2}}$$

$$\varphi_1 = \tan^{-1}\frac{2\delta\Omega}{\omega_0^2 - \Omega^2}$$

From the equation

$$\frac{dA}{d\Omega} = 0$$

it follows that the extreme value of the frequency Ω_0 of the driving force F_0, at which the amplitude of forced oscillations takes a maximum value, is

$$\Omega_0 = \sqrt{\omega_0^2 - 2\delta^2} \qquad (5.61)$$

Maximum amplitude corresponding to frequency Ω_0

$$A_{max} = \frac{F_0}{2L\delta\omega_0} \qquad (5.62)$$

A comparison of (5.62) and (5.47) shows that the frequency Ω_0 is slightly less than the natural frequency of the dissipative system.

The phenomenon of a sharp increase in the amplitude of forced oscillations of the system, when the cyclic frequency of the driving force approaches a value roughly equal to the natural frequency of the system, is called **resonance, and the frequencyΩ_0 is called the resonance frequency**.

The curves of dependence of the amplitude and phase of forced oscillations for different values of δ on the cyclic frequency Ω of the driving force are called **resonance curves** (Fig. 5.4 a and b).

If a source of emf is included in an oscillating circuit, the value of which varies according to the harmonic law

$$E_C = E_{C0} \sin \Omega_t \tag{5.63}$$

then the forced oscillations of frequency Ω will occur in the circuit.

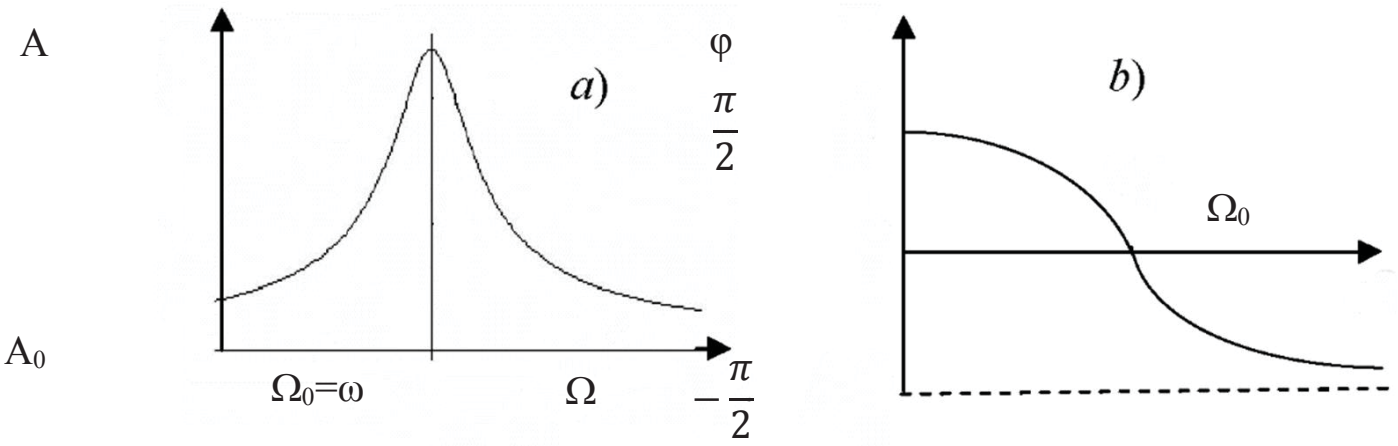


Figure 5.4.

The current in the circuit can be found from the relation

$$I = \frac{dq}{dt} \tag{5.64}$$

Substitution gives for the current amplitude I

$$I_0 = E_{C0} \frac{1}{\sqrt{R^2 + (\frac{1}{\Omega C} - \Omega L)^2}} \tag{5.65}$$

Where

$$Z = \sqrt{R^2 + (\frac{1}{\Omega C} - \Omega L)^2} \tag{5.66}$$

is the value of the total circuit resistance, called **impedance**.

At resonance, according to (5.62) and (5.64),

$$I_0 = E_{C0}\frac{1}{R} \tag{5.67}$$

This means that under conditions of resonance, the amplitude of the current in the circuit is as large as possible, and

$$\frac{1}{\Omega C} = \Omega L \tag{5.68}$$

Since $\Omega \approx \omega_0 - \varphi 0$, (5.68) takes the form $X_C = X_L$.

5.1.4. Vector form of harmonic oscillations

Let a harmonic oscillation be given in the form

$$q = A\cos(\omega t + \varphi_0) \tag{5.69}$$

From an arbitrarily chosen point O in the plane, at an angle φ_0 to the horizontal axis Oq, we lay out a vector $\vec{OA}$ whose modulus is equal to the amplitude of the vibrations A. (Fig. 5.5)

The obtained vector $\vec{OA}$ is rotated with angular velocity ω counterclockwise. In time t, the end of the vector $\vec{OA}$ will make an angle equal to ωt and take the position of OA_1 so that at time t it forms the following angle with the horizontal axis

$$\varphi = \omega t + \varphi_0 \tag{5.70}$$

that is equal to the phase of the oscillations. As can be seen in Fig. 5.5, the projection of vector $\vec{OA}_1$ on the horizontal axis is

$$q(t) = OA\cos\varphi = A\cos(\omega t + \varphi_0) \tag{5.71}$$

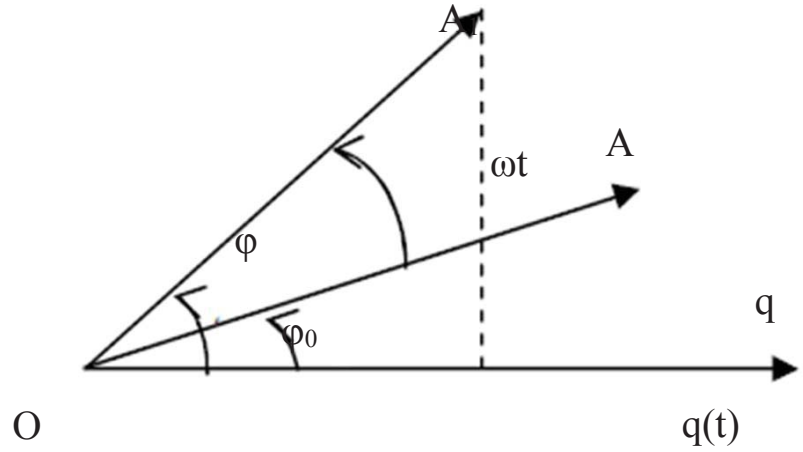


Figure 5.5.

Thus, as the vector $O\vec{A}$ rotates, its projection on the horizontal axis oscillates according to a given initial harmonic law.

From this we conclude that a harmonic oscillation in the general case can be given graphically on the plane in the form of a vector, whose length is equal to the amplitude of oscillations, and its direction at a given time forms with the horizontal axis an angle equal to the phase of oscillations.

Such a representation of harmonic oscillations in the form of vectors rotating with cyclic frequency (Fig.5.5) is called a **vector diagram**.

5.1.5. Addition of harmonic oscillations

Let us assume that it is necessary to combine two harmonic oscillations x_1 and x_2 made in the same direction along the horizontal axis Ox with frequency ω, with amplitudes A_1 and A_2 and initial phases φ_{01} and φ_{02},

$$\left.\begin{aligned} x_1 &= A_1\cos(\omega t + \varphi_{01}) \\ x_2 &= A_2\cos(\omega t + \varphi_{02}) \end{aligned}\right\} \qquad (5.72)$$

Let's represent these oscillations with the vector diagram of vectors $O\vec{A}_1$ and $O\vec{A}_2$ (Fig. 5.6).

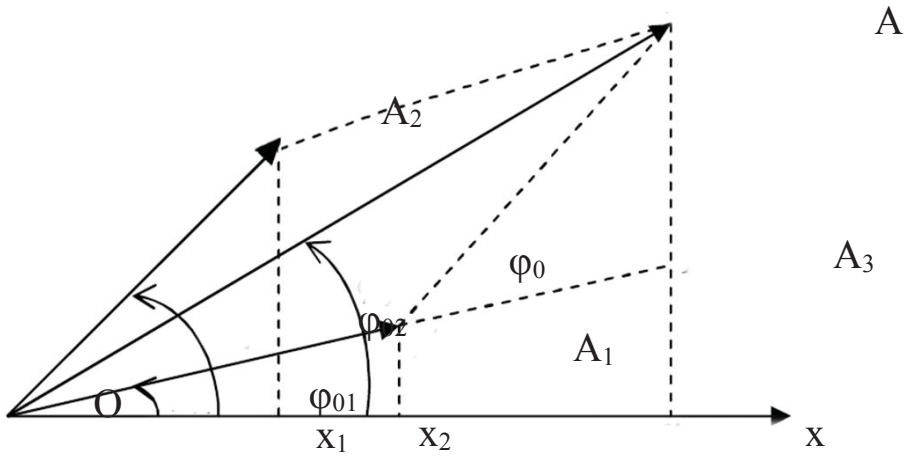


Figure 5.6

Since both oscillations occur with the same cyclic frequency, when added together they are represented by vectors inclined to the horizontal axis at angles φ_{01} and φ_{02}.

The resulting vector $\vec{OA}$ is then rotated by the angle ωt (not shown in the figure). The vector diagram shows that

$$x = x_1 + x_2 = A_1\cos(\omega t + \varphi_{01}) + A_2\cos(\omega t + \varphi_{02}) = A\cos(\omega t + \varphi_0)$$

The values of A and φ_0 of the resulting oscillation can be determined from the vector diagram. The direction of the vector $\vec{OA}$ is determined by the parallelogram rule (see Fig. 5.6), and its length by the cosine theorem for the triangle OAA_1 (Appendix 1).

$$OA^2 = OA_1^2 + OA_2^2 - 2OA_1 \cdot OA_2\cos\angle OA_1A \qquad (5.73)$$

($AA_1 = OA_2$ as opposite sides of a parallelogram). As can be seen from Fig. 5.6,

$$\angle OA_1A = \pi - \angle AA_1A_3,$$

$$\angle AA_1A_3 = \angle(\varphi_{02} - \varphi_{01})$$

as angles with mutually parallel sides. Therefore

$$OA^2 = OA_1^2 + OA_2^2 - 2OA_1OA_2\cos[\pi - (\varphi_{02} - \varphi_{01})] \qquad (5.74)$$

According to the formulas for the reduction

$$\cos(\pi - \alpha) = - \cos\alpha.$$

Thus, it follows from (5.74) that

$$A^2 = A_1^2 + A_2^2 + 2A_1A_2 \cos(\varphi_{02} - \varphi_{01}) \qquad (5.75)$$

Angle φ_0 can be found from the triangle OAx, namely

$$\tan \varphi_0 = \frac{Ax}{Ox} \qquad (5.76)$$

Fig. 5.6 shows that

$$\left.\begin{aligned} Ax &= A_1x_1 + A_2x_2 \\ Ox &= x_1 + x_2 \end{aligned}\right\} \qquad (5.77)$$

On the other side,

$$\left.\begin{aligned} A_1x_1 &= A_1 \sin \varphi_{01} \ ; x_1 = A_1 \cos \varphi_{01} \\ A_2x_2 &= A_2 \sin \varphi_{02} \ ; x_2 = A_2 \cos \varphi_{02} \end{aligned}\right\} \qquad (5.78)$$

Substituting (5.78) into (5.76) gives

$$\tan \varphi_0 = \frac{A_1 \sin \varphi_{01} + A_2 \sin \varphi_{02}}{A_1 \cos \varphi_{01} + A_2 \cos \varphi_{02}} \qquad (5.79)$$

Let us consider, using relations (5.75) and (5.79), particular cases of addition of unidirectional harmonic oscillations. If the phase difference of such oscillations

$$\Delta\varphi = \varphi_{02} - \varphi_{01} = 2\pi k, \text{ where } k = 0,1,2...,$$

then because

$$\cos 2\pi k = 1,$$

it follows from (5.75) that

$$A^2=(A_1 + A_2)^2$$

and

$$A=A_1 + A_2 \qquad (5.80)$$

If the phase difference

then

$$A^2=(A_1 + A_2)^2$$

and

$$A = |A_2 – A_1| \qquad (5.81)$$

It follows from (5.80) and (5.81) that in the first case the oscillations, adding up, strengthen, and in the second case they, on the contrary, mutually weaken.

When an oscillation is non-harmonic, it can, as mentioned above, be represented as a sum of harmonic oscillations of multiple frequency (Fourier series). Each of these oscillations can also be represented by vectors that rotate at different angular velocities.

Harmonic oscillations with different angular velocity can also be added using a vector diagram. Each such oscillation at a given time can be associated with a vector, the direction of which in the general case is a function of time. The amplitude of the resulting oscillation is also a function of time.

Moreover, the time also determines the magnitude of

$$\Delta\varphi = \varphi_{02} - \varphi_{01}$$

The resulting vector in this case will rotate with a variable speed. For these reasons, it is a non-harmonic oscillation. Finding the resultant oscillation in the general case is associated with serious mathematical difficulties. Such an oscillation cannot always be represented by a simple enough formula. Let us consider a simpler special case of addition of two harmonic oscillations with the same amplitudes and initial phases equal to zero and different, but quite close, frequencies ω and ω+Δω. In the case where

$$\Delta\omega \ll \omega \qquad (5.82)$$

these oscillations can be written in the form

$$\left.\begin{aligned} x_1 &= A\cos\omega t \\ x_2 &= A\cos(\omega + \Delta\omega) \end{aligned}\right\} \qquad (5.83)$$

Let's add x_1 *and* x_2 analytically, without resorting to a vector diagram

$$x = x_1 + x_2 = A\cos\omega t + A\cos(\omega + \Delta\omega)t = A[\cos\omega t + \cos(\omega + \Delta\omega)t]$$

Let us apply the formula known from trigonometry for the addition of cosines

$$x = 2A\cos\frac{\Delta\omega}{2}t\cos\left(\omega + \frac{\Delta\omega}{2}\right)t \qquad (5.84)$$

Taking into account (5.82), the total oscillation is finally

$$x = 2A\cos\frac{\Delta\omega}{2}t\cos\omega t \qquad (5.85)$$

The resulting oscillation can be conventionally considered harmonic with a variable amplitude changing according to the harmonic law with a frequency Δω /2.

Equation (5.85) describes the phenomenon of so-called **beatings** of two oscillations close in frequency. Physically, beatings are explained by the fact that one oscillation is constantly lagging behind the other, and the total amplitude increases when the phases coincide, and decreases when they do not coincide.

The oscillation graph at beating is shown in Fig. 5.7.

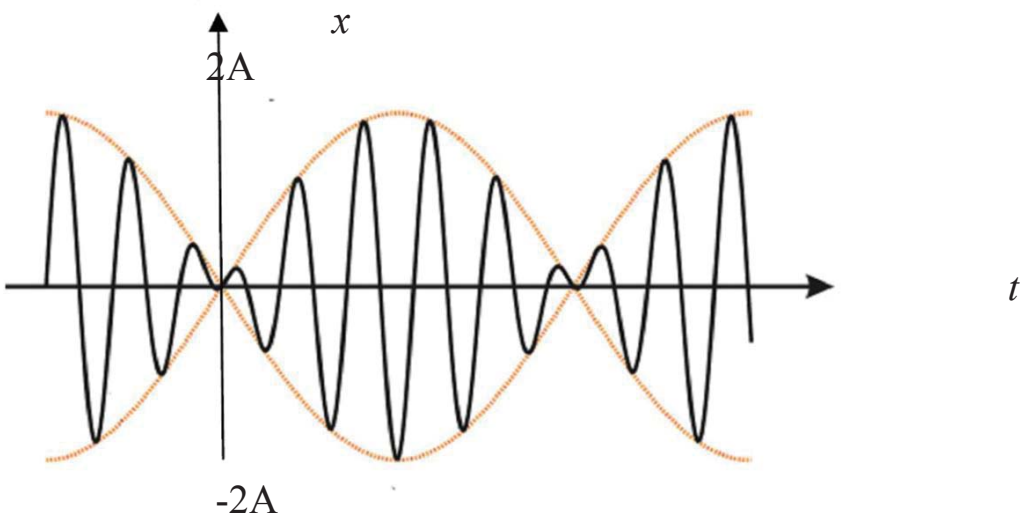


Figure 5.7

* Let us now consider the rule of addition of harmonic oscillations occurring in different directions. Suppose, for example, it is necessary to add two oscillations of the same frequency, which are made in mutually perpendicular directions. Let us assume that the oscillations occur along the Ox and Oy axes of the Cartesian coordinate system. Let us assume that the phase shift of the oscillations $\Delta\varphi = \varphi_0$. At the same time

$$x = A\cos\omega t\,, y = B\cos(\omega t + \varphi_0) \qquad (5.87)$$

From the first equation we find

$$\cos\omega t = \frac{x}{A},$$

therefore,

$$\sin\omega t = \pm\sqrt{1 - \frac{x^2}{B^2}}$$

Let us further represent the second equation (5.87) taking into account the rules of trigonometry in the form

$$y = B\cos(\omega t + \varphi_0) = B[\cos\omega t\cos\varphi_0 - \sin\omega t\sin\varphi_0],$$

$$y = B\left[\frac{x}{B}\cos\varphi_0 \pm \sqrt{1-\frac{x^2}{A^2}}\,\sin\varphi_0\right] \tag{5.88}$$

from which

$$\frac{y}{B} = \frac{x}{A}\cos\varphi_0 \pm \sqrt{1-\frac{x^2}{A^2}}\sin\varphi_0$$

or

$$\frac{y^2}{B^2} = \frac{x^2}{A^2}\cos^2\varphi_0 + \left(1-\frac{x^2}{A^2}\right)\sin^2\varphi_0 \pm 2\frac{x}{A}\sqrt{1-\frac{x^2}{A^2}}\sin\varphi_0\cos\varphi_0 =$$

$$= \frac{x^2}{A^2}\cos 2\varphi_0 + \frac{x}{A}\sqrt{1-\frac{x^2}{A^2}}\sin 2\varphi_0 + \sin^2\varphi_0$$

Let's add to both parts $\frac{x^2}{A^2}$, then

$$\frac{y^2}{B^2} + \frac{x^2}{A^2} = \frac{x^2}{A^2}(1+\cos 2\varphi_0) + \frac{x}{A}\sqrt{1-\frac{x^2}{A^2}}\sin 2\varphi_0 + \sin^2\varphi_0 =$$

$$= 2\frac{x}{A}\cos\varphi_0\left(\frac{x}{A}\cos\varphi_0 + \sqrt{1-\frac{x^2}{A^2}}\,\sin\varphi_0\right) + \sin^2\varphi_0$$

Since the expression in parentheses is equal to $\frac{y}{B}$, (see 5.88), then

$$\frac{y^2}{B^2} + \frac{x^2}{A^2} - 2\frac{xy}{AB}\,\cos\varphi_0 = \sin^2\varphi_0 \tag{5.89}$$

Equation (5.89) describes the trajectory of the resultant oscillation point.

In the case where

$\varphi_0 = n\pi$, *where n = 0, 1, 2, ...*

equation (5.89) takes the form

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} - (-1)^n \frac{2xy}{AB} = 0 \qquad (5.90)$$

For even *n (n = 2 k),* we obtain

$$\frac{x}{A} - \frac{y}{B} = 0$$

or

$$y = \frac{B}{A}x \qquad (5.91)$$

At odd *n*

$$\frac{x}{A} + \frac{y}{B} = 0$$

or

$$y = -\frac{B}{A}x \qquad (5.92)$$

Equations (5.91) and (5.92) are equations of lines passing through the origin at an angle ± π / 4. In the case where

$$\varphi_0 = \pm(1 + 2\pi)\frac{\pi}{2},$$

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} = 1 \qquad (5.93)$$

Equation (5.93) is the canonical equation of an ellipse symmetric about the origin (Fig. 5.8a). In all other cases, equation (5.89) is also the equation of a family of ellipses whose semi-axes are inclined to the coordinate axes at different angles determined by the angle φ_0 (see Fig. 5.8 b). If A=B, the trajectories of ellipses degenerate into circles. When mutually perpendicular oscillations occur with different frequencies, the trajectory of the resulting oscillation point has a rather complex form.

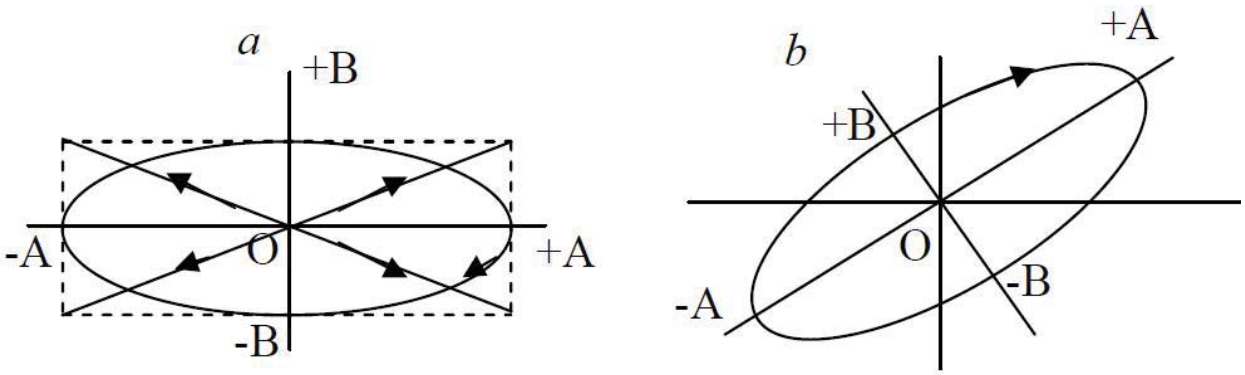


Figure 5.8.

The trajectory of the point, describing the resulting oscillation, for example, in the case when the frequencies of the folded oscillations are multiples of 2, has the form shown in Fig. 5.9 (accepted without proof)

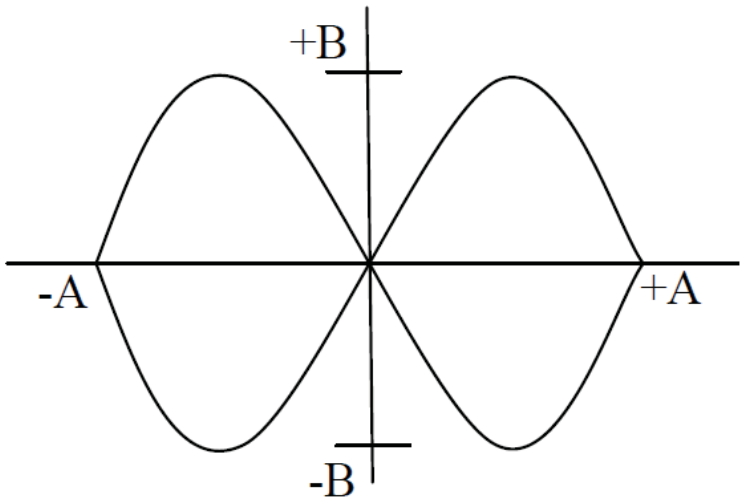


Figure 5.9.

The trajectories of the point describing the resulting oscillations are called **Lissajous** curves. Lissajous curves can be obtained on the oscilloscope screen by applying electric signals, corresponding to the added oscillations, to the opposite deflecting plates.

5.1.6. Linear oscillatory systems

When considering harmonic oscillations, it is advisable to use the notion of phase space and phase trajectories introduced in Section 2.5. In the case of oscillating systems with one degree of freedom, the phase space turns into a phase plane. In this plane, in the Cartesian coordinate

system, the value of the dynamic variable characterizing the oscillating system, for example q(t), is plotted along the horizontal axis and its time derivative is plotted along the ordinate axis $\dot{q}(t)$. In this case, the evolution of the system is described by a phase trajectory $\dot{q} = f[q(t)]$

The geometric picture of the mutual arrangement of phase trajectories on the phase plane forms the phase portrait of the oscillatory process.

Let us first consider the phase portrait of free (natural) undamped oscillations, which, as mentioned above, are described by the equations of motion of the following form

$$\ddot{m}x = -kx \tag{5.94}$$

Let us multiply both parts of the equation by $\dot{x}$.

Then

$$\frac{d}{dt}\left[\frac{m}{w}\dot{x}^2\right] + \frac{d}{dt}\left[\frac{k}{w2}x^2\right] = 0 \tag{5.95}$$

By integrating the equation (5.95), we obtain

$$\frac{m}{2}\dot{x^2} + \frac{k}{2}x^2 = cosnt \tag{5.96}$$

Since the left part of equation (5.96) contains the sum of kinetic and potential energy of the oscillating system, the integration constant in the right part of the equation has the meaning of the total energy of the system ε, i.e.

$$\frac{mx^2}{2} + \frac{kx^2}{2} = \varepsilon \tag{5.97}$$

Let us denote the quantity $\dot{x}$ by *y*, then

$$\frac{my^2}{2\varepsilon} + \frac{kx^2}{2\varepsilon} = 1 \tag{5.98}$$

Let us introduce the values a and b such that

$$\frac{2\varepsilon^2}{k} = a^2; \frac{2\varepsilon}{m} = b^2$$

then the equation (5.89) takes the form

$$\frac{y^2}{b^2} + \frac{x^2}{a^2} = 1 \qquad (5.99)$$

Thus, the equation of the oscillatory process is geometrically expressed by the equation of an ellipse. Varying the value of energy $\boldsymbol{\varepsilon}$ leads to a family of phase trajectories, that is, the phase portrait of the considered process (Fig. 5.10)

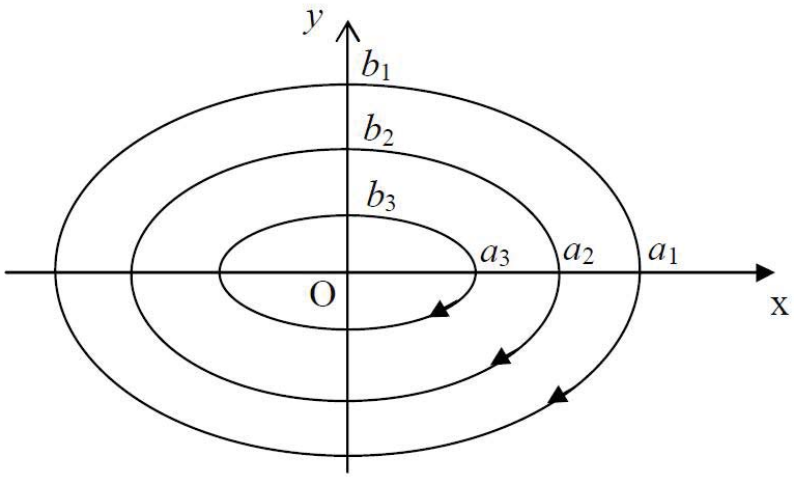


Figure 5.10.

It is not hard to see that

$$\frac{b_i}{a_i} = \sqrt{\frac{k}{m}} = \omega_0$$

where b_i *and* a_i are the semi-axes of the family of ellipses.

The arrows on the phase trajectories indicate the direction of motion of the current point $[x(t), \dot{x}(t)]$ over time. This point is called the **representation point**.

It can be shown (we will accept this without proof) that the phase trajectories of all oscillatory (periodic) processes are always closed curves.

This rule allows us to distinguish periodic processes from aperiodic ones. In particular, it is used to predict the behavior of the system by the type of the differential equation of motion describing it and, accordingly, the phase space and phase portrait.

The point with coordinates (0,0) is a special point. It characterizes the equilibrium state of the system and can be regarded as a phase trajectory of the equilibrium state.

The special point in the vicinity of which all phase trajectories form closed lines is called the **center**. In this sense, the equilibrium state of the system performing harmonic oscillations is a special point of the "Center" type.

Let us divide both parts of equation (5.97) by $\omega_0^2\varepsilon$ and denote

$$y = \frac{\dot{x}}{\omega_0}$$

then

$$\frac{y^2}{\frac{2\varepsilon}{m}} + \frac{x^2}{\frac{2\omega_0^2\varepsilon}{k}} = \frac{1}{\omega_0^2} \qquad (5.100)$$

Since

$$2\omega_0^2\frac{\varepsilon}{k} = 2\frac{k}{m}\frac{\varepsilon}{k} = \frac{2\varepsilon}{km}$$

then when denoting

$$A^2 = \frac{2\varepsilon}{k}$$

We will get

$$y^2 + x^2 = A^2 \qquad (5.101)$$

Since

$$\varepsilon = \frac{kA^2}{2}$$

a is the amplitude of oscillations, then equation (5.101) depicts a family of circles with radius $A = \sqrt{\frac{2\varepsilon}{k}}$

This means that oscillations with the same amplitudes but different phases correspond to the motion of the representation point on the same circle of radius A. At any given moment this circle represents points 1 and 2 of two oscillations with the same amplitude and phase shift $\Delta\omega=\varphi_2-\varphi_1$, and the angle of phase shift $\Delta\omega$ is equal to the angle between the position vectors of points 1 and 2

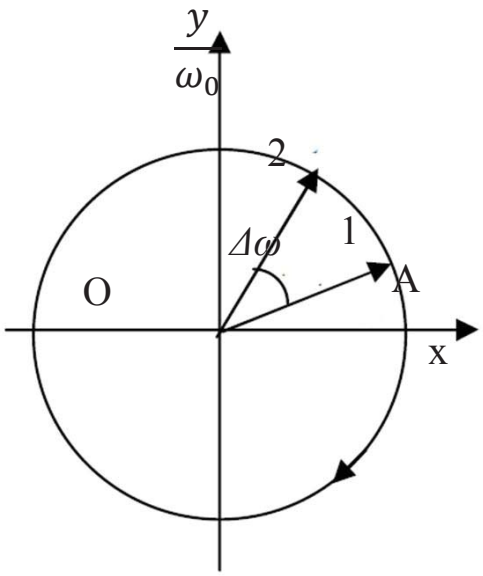


Figure 5.11

In the case of a damped oscillatory process, it is quite difficult to obtain an equation describing the phase trajectory. Therefore, let us give examples of phase trajectories of an aperiodic process (Fig. 5.12a) and a periodic process (Fig. 5.12b) without their calculation. As can be seen from Fig. 5.12, the phase trajectories of these processes are depicted by open lines

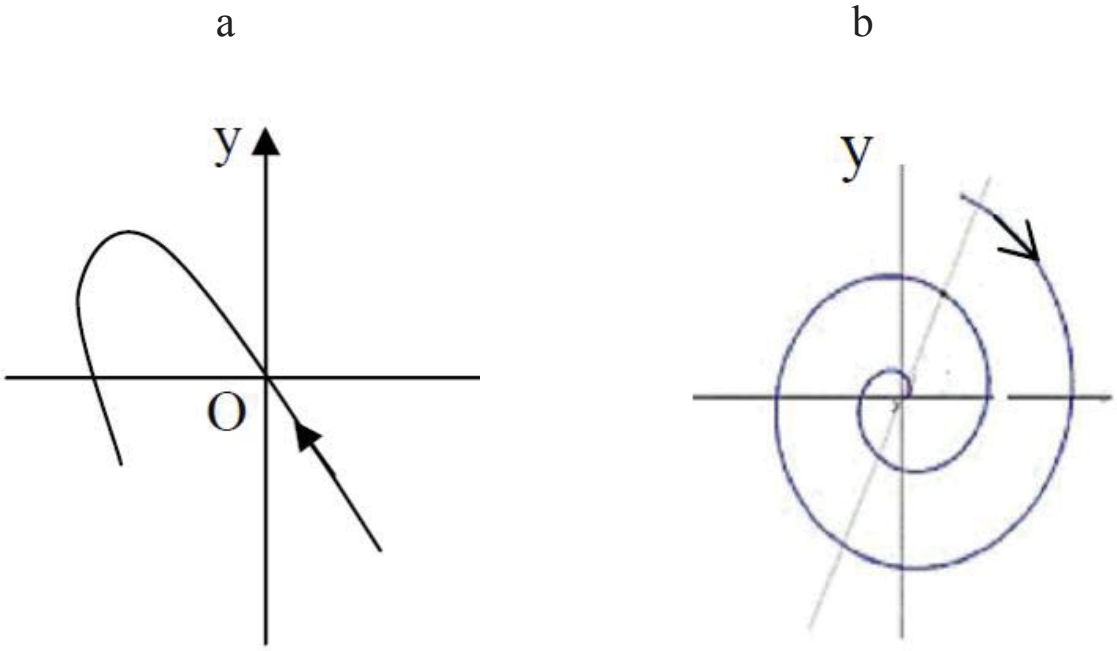


Figure 5.12.

5.1.7. Non-linear oscillatory systems

Linear oscillatory processes are an idealization of the actual reality and approximate it under conditions of small perturbations. Linear processes are a consequence of the superposition principle, according to which it is assumed that if several forces act on a system, then the behavior of the system under the action of each force does not depend on the action of other forces, and the result of the behavior of the system under the action of other forces is obtained by simple addition of the results of each force. In reality, on the contrary, the behavior of one element of a system under the action of a given force significantly affects the behavior of other elements. In this case, the behavior of the whole system is described by nonlinear differential equations.

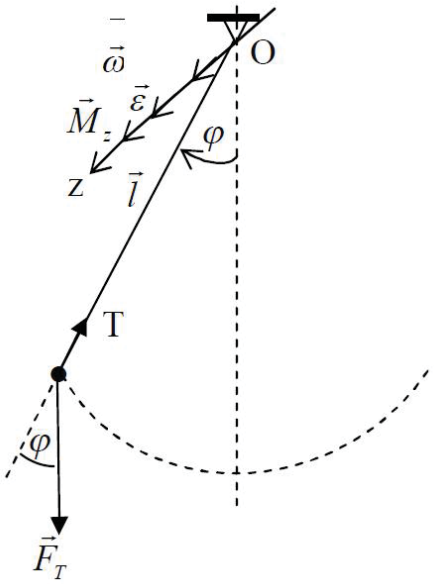


Figure 5.13.

Nonlinear oscillatory systems perform rather complex non-harmonic oscillations. These oscillations occur even in the simplest system of a mathematical pendulum deflected from equilibrium by a large enough angle. Suppose, for example, a mathematical pendulum of length l with mass m (Fig. 5.13). Let us bring it out of the equilibrium position by deflecting it by a sufficiently large angle φ, which will cause it to move along an arc of a circle around the axis Oz. The following symbols are used in the figure:

$\vec{F}_T = m\vec{g}$ is the force of gravity acting on the pendulum (load);

$\vec{T}$ is the force of the tension of the pendulum string;

$\vec{\omega}$ is the vector of the angular velocity of the load;

$\vec{\varepsilon}$ is the acceleration vector;

$\overrightarrow{M_z}$ is the vector of the momentum of gravity of the load with respect to the axis Oz

It was shown above that the equation of rotational motion of an absolutely solid body (see Section 2.7.10), in this case a ball of a mathematical pendulum, has the form -*

$$\vec{M}_z = I_z\vec{\varepsilon} \qquad (5.102)$$

where I_z is the moment of inertia of the ball as it moves around the axis.

According to the definition, the moment of force is

$$M_z = \left[\vec{l}, m\vec{g}\right] \tag{5.103}$$

from which

$$\vec{M}_z = mgl \sin \varphi$$

Further

$$\omega = \frac{d\omega}{dt}\vec{k} \tag{5.104}$$

where $\vec{k}$ is the unit vector of the Oz axis.

Therefore

$$\vec{\varepsilon} = \frac{d^2\varphi}{dt^2}\vec{k} \tag{5.105}$$

$$I_z = ml^2 \tag{5.106}$$

Projecting equation (5.102) onto the Oz axis gives

$$mgl \sin \varphi = -ml^2 \frac{d^2\varphi}{dt^2} \tag{5.107}$$

from which

$$\ddot{\varphi} + \frac{g}{l} \sin \varphi = 0 \tag{5.108}$$

Since for a mathematical pendulum

$$\frac{g}{l} = \omega_0^2,$$

then finally the equation of oscillations of the pendulum

$$\ddot{\varphi} + \omega_0^2 \sin \varphi = 0 \tag{5.109}$$

(5.109) is a second-order nonlinear differential equation in which the dependence $sin\varphi = f(t)$ requires additional conditions. There is,

however, a simpler approximate method, which consists in decomposing the function *sinφ* into a series of powers of φ and limiting by some criteria the number of members of this series.

In this case

$$\sin\varphi \cong \varphi - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \cdots + (-1)^n \frac{\varphi^{2n+1}}{(2n+1)!} \tag{5.110}$$

Limiting ourselves to two terms of a series with relatively small values of φ, we can write equation (5.109) in the form

$$\ddot{\varphi} + \omega_0^2 \left(\varphi - \frac{\varphi^3}{6}\right) \approx 0 \tag{5.111}$$

The nonlinear equation (5.111) is solvable in principle, although its solution is associated with great mathematical difficulties. It can, in particular, be solved by the method of successive approximations. As a first approximation, let us assume that the third term of the equation can be neglected, then we obtain

$$\varphi_1 + \omega_0^2 \varphi_1 \approx 0$$

The solution to this linear equation is a well-known harmonic function

$$\varphi_1 = \varphi_0 \sin\omega t \tag{5.112}$$

where φ_0 is the amplitude of oscillations.

In the second approximation, we look for the solution of equation (5.111) in the form $\varphi_2 \approx \varphi_1$, then

$$-\omega^2 \varphi_0 \sin\omega t + \omega_0^2 \left(\varphi_0 \sin\omega t - \frac{\varphi_0^3 \sin^3 \omega t}{6}\right) = 0$$

Since

$$\sin^3 \omega t = -\frac{3}{4}\sin\omega t - \frac{1}{4}\sin 3\omega t$$

then

$$\left(\omega_0^2 - \omega^2 - \frac{\omega_0^2\varphi_0^2}{8}\right)\varphi_0 \sin\omega t + \omega_0^2 \frac{\varphi_0^3}{8}\sin 3\omega t \cong 0$$

This equality, however, is not realized under any conditions due to the presence of an odd harmonic $sin3\omega t$, so in the third approximation we look for a solution in the form

$$\varphi_3 \cong \varphi_0 \sin\omega t + \varepsilon\varphi_0 \sin 3\omega t \qquad (5.113)$$

However, cubing p3 leads to the appearance of odd harmonics of even higher order, etc. By continuing to adjust the solution with higher and higher approximations, we finally obtain a solution that is close to the true solution. In doing so, it turns out that

$$\omega \approx \omega_0(1 - \frac{\varphi_0^2}{16}) \qquad (5.114)$$

It follows from (5.114) that the frequency of oscillations of a nonlinear system is less than the natural frequency of a linear system and, in addition, depends not only on the properties of the system, but also on the amplitude of oscillations φ_0. The greater the amplitude, the lower the frequency. From the approximate solution (5.114) we see that it is non-harmonic or, as they say, **anharmonic**.

Finally, it is quite obvious that the principle of superposition loses its meaning for nonlinear oscillations.

Another practically very important example of nonlinear oscillations is self-oscillations. The system itself, in which self-oscillations occur, is dissipative, i.e. there are continuous losses of oscillatory energy. Meanwhile, despite the autonomy of the system, it maintains undamped oscillations in automatic mode due to the built-in source of energy, which compensates losses.

Obviously, the energy supply to the system should not be constant, but synchronized with the cycle of oscillations. In other words, the embedded energy source must also be an oscillating system with oscillations of the same frequency as the main system. This is achieved

by means of a so-called feedback device with an appropriate valve that synchronizes the power supply.

Examples of self-oscillating systems are sinusoidal signal generators, mechanical clocks, internal combustion engines, the heart of biological systems and many others.

For example, in mechanical watches, the built-in source of energy is a spiral spring, the valve is an anchor mechanism, and the feedback is provided through a system of gears. In a sinusoidal signal generator the built-in energy source is an electric battery, the feedback is provided by inductances or capacitances, and the valve is an appropriate semiconductor device (e.g. thyristor), etc.

5.1.8. Complex form of harmonic oscillations

As is known (Appendix 1), if in the Cartesian coordinate system we plot real numbers on the horizontal axis Ox and imaginary numbers on the vertical axis, any complex number

$$\hat{c} = a + jb$$

can be represented by a point in the plane with coordinates a and b, equal respectively to the real and imaginary parts of the complex number.

In this case

$$a = Re\hat{c}; b = In\hat{c}$$

On the other hand, any point in the plane (i.e. a given complex number) can be matched with a vector $0\vec{A}$, coming from the origin (Fig. 5.14).

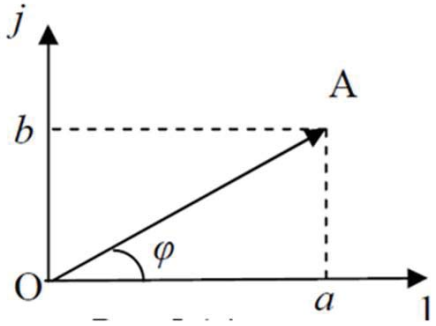


Figure 5.14.

Point A in the picture represents the complex number $\hat{c} = a + jb$. As can be seen from the figure the vector $\vec{OA}$, coordinates are respectively

$$\left.\begin{aligned} a &= OA\cos\varphi \\ b &= OA\sin\varphi \end{aligned}\right\} \tag{5.115}$$

It is, in this way,

$$\hat{c} = OA(\cos\varphi + j\sin\varphi) \tag{5.116}$$

On the other hand, according to the Euler formula (Appendix 2)

$$\cos\varphi + j\sin\varphi = e^{j\varphi} \tag{5.117}$$

where $e^{j\varphi}$ is the complex exponent.

It is, in this way,

$$\hat{c} = OAe^{j\varphi}$$

where OA is the vector length

$$OA = \sqrt{a^2 + b^2},$$

and the angle φ is defined (see Fig. 5.14) by the formula

$$\varphi = \tan^{-1}\frac{b}{a} \tag{5.118}$$

From (5.117) we see that the modulus of the complex number

$$\left|e^{j\varphi}\right| = \sqrt{cos^2\varphi + sin^2\varphi} = 1$$

Since a harmonic oscillation has the form

$$q = A\cos(\omega t + \varphi_0),$$

then it can always be treated as the real part of the complex number $\hat{c}$ for which the amplitude is equal to the absolute value $|\hat{c}| = OA$, and the phase

$$\omega t + \varphi_0 = \varphi.$$

In other words, if a harmonic oscillation is given

$$q = Acos(\omega t + \varphi_0),$$

then in the complex form it has the following form

$$q = ReAe^{j(\omega t + \varphi 0)}$$

Since in the vector representation of the oscillation only the initial phase is taken into account, usually an oscillation of the form

$$q = Acos(\omega t + \varphi_0)$$

is presented in complex form as

$$\hat{q} = Ae^{j\varphi_0}$$

and as a result

$$q = Re(\hat{q} \cdot e^{j\varphi t}) \tag{5.119}$$

Let us apply a complex representation of harmonic oscillations to the study of forced oscillations. To do this, we turn from the value of q to the complex quantity of $\hat{q}$. When obtaining the final result of the calculations, we return to the original non-complex quantities by formula (5.119). Let's write down the initial equation of forced oscillations in the complex form

$$\ddot{\hat{q}} + 2\delta\dot{\hat{q}} + \omega_0^2\hat{q} = \frac{F_0}{L}e^{j\Omega t} \tag{5.120}$$

We find its solution as the sum of the solutions of the homogeneous equation and the partial non-homogeneous one. In the steady-state mode, the natural oscillations are damped and do not affect the overall result. Therefore, the desired solution is reduced only to finding a partial solution of an non-homogeneous equation. We look for it as the right part of (5.120) in the complex form, namely

$$\hat{q}_2 = \hat{A}e^{j\Omega t} \tag{5.121}$$

where $\hat{A}$ is the complex amplitude.

* In the general case

$$\hat{A} = \rho e^{j\varphi} \tag{5.122}$$

where ρ is the absolute value of the complex number $\hat{q}_2$

It follows from (5.121) that

$$\left.\begin{aligned}\dot{\hat{q}}_2 &= \hat{A}j\Omega e^{j\Omega t}\\ \ddot{\hat{q}}_2 &= \hat{A}j\Omega^2 e^{j\Omega t}\end{aligned}\right\}$$

Substitution gives the expression

$$-\hat{A}\Omega^2 e^{j\Omega t} + 2\delta\hat{A}j\Omega e^{j\Omega t} + \omega_0^2\hat{A}e^{j\Omega t} = \frac{F_0}{L}e^{j\Omega t} \qquad (5.123)$$

Reducing (5.123) by the common factor $e^{j\Omega t}$, we obtain that

$$A = \frac{F_0}{[(\omega_0^2 - \Omega^2) + 2j\delta\Omega]L} \qquad (5.124)$$

Let us express (5.124) as a complex quantity

$$\hat{A} = a + jb = Re\hat{A} + jIm\hat{A}$$

To do this, multiply the numerator and denominator of expression (5.124) by the complex conjugate of the denominator

$$[(\omega_0^2 - \Omega^2) + 2j\delta\Omega],$$

then

$$\hat{A} = \frac{F_0[(\omega_0^2 - \Omega^2) + 2j\delta\Omega)]}{[(\omega_0^2 - \Omega^2) + 4j\delta^2\Omega^2]L}$$

from which

$$Re\hat{A} = \frac{F_0(\omega_0^2 - \Omega^2)}{L[(\omega_0^2 - \Omega^2)^2 + 4j\delta^2\Omega^2]}; Im\hat{A} = -\frac{2F_0\delta\Omega}{L[(\omega_0^2 - \Omega^2)^2 + 4j\delta^2\Omega^2]}$$

It is, in this way,

$$\rho = \sqrt{(ReA)^2 + (ImA)^2} = \frac{F_0}{L\sqrt{(\omega_0^2 - \Omega^2)^2 + +4j\delta^2\Omega^2}} \qquad (5.125)$$

$$\varphi = -\tan^{-1}\frac{2\delta\Omega}{\omega_0^2 - \Omega^2} \qquad (5.126)$$

Finally

$$q_2 = \frac{\frac{F_0}{L}}{\sqrt{(\omega_0^2 - \Omega^2)^2 + 4\delta^2\Omega^2}} \cos\left[\Omega t + \tan^{-1}\frac{2\delta\Omega}{\omega_0^2 - \Omega^2}\right] \tag{5.127}$$

For electromagnetic oscillations in a circuit with external emf $Ec_0 \sin\Omega t$

$$\frac{F_0}{L} = Ec_0;\ \rho = I_0$$

after substituting the values ω_0 и δ

$$I_0 = Ec_0 \frac{1}{\sqrt{R^2 + \left(\Omega L - \frac{1}{\Omega C}\right)^2}} \tag{5.128}$$

where, as has been shown, the impedance Z is

$$Z = \sqrt{R^2 + \left(\Omega L - \frac{1}{\Omega C}\right)^2} \tag{5.129}$$

or

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \tag{5.130}$$

Let us transform (5.124) and write it in the form

$$\hat{A} = \frac{F_0}{j\Omega\left[\left(j\Omega\mathrm{L} - \frac{1}{\Omega\mathrm{C}}\right) + R\right]}$$

It is easy to see that the value in square brackets has the dimension of resistance and can be regarded as a complex impedance

$$\hat{Z} = R + j(X_L - X_C) \tag{5.131}$$

In other words, the reactive resistances in complex form have the form

$$\hat{X}_L = j\Omega \mathrm{L} = \Omega \mathrm{L} e^{+j\frac{\pi}{2}} \tag{5.132}$$

$$\hat{X}_C = j\frac{1}{\Omega \mathrm{C}} = \frac{1}{\Omega \mathrm{C}} e^{-j\frac{\pi}{2}} \tag{5.133}$$

5.2. Alternating electric current

If we include a source of periodically varying EMF into the oscillating circuit (Fig. 5.3) as a battery, then there will be forced oscillations of alternating current with the frequency of the source in the RLC-loop of the circuit. Let's look at calculations of alternating current circuits with different elements.

5.2.1. Calculation of a circuit with active resistance

Figure 5.15 shows an alternating current circuit consisting of an active resistance R and a source of harmonic emf $E_C(t)$, obeying the law

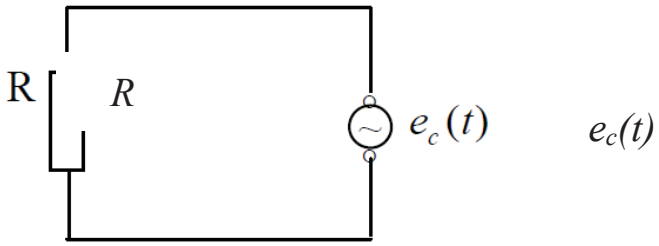


Figure 5.15.

$$e_c(t) = Ec_0 cos\Omega t,$$

where Ω is the frequency of the generated emf;

Ec_0 is the amplitude of the emf.

Under the action of this emf of an electric circuit, an alternating voltage is created

$$u(t) = U_m \cos \Omega t \qquad (5.134)$$

As can be seen from relation (5.128), under the action of voltage *u(t)* under the condition that $X_L = X_C = 0$ and $R \neq 0$, the current

$$i(t) = \frac{U_m cos\Omega t}{R} \qquad (5.134)$$

where R is the active resistance of the circuit, equal to the real part of the complex resistance (see 5.131).

In other words, in this case

$$Z = R. \qquad (5.135)$$

Equation (5.134) can be regarded as Ohm's law for a circuit section. As can be seen from (5.134) the active resistance does not change the phase of alternating current oscillations in the circuit, and therefore the phase shift between current and voltage on the resistance is zero.

From Fig. 5.15 and the relations for current and voltage, it follows that all the power in the circuit in question is allocated on the active resistance. Herein lie both the advantages and disadvantages of this circuit. The advantage is that all the power is used by the consumer. The disadvantage is that on the resistance, in addition to the useful power, joule heat is also released, which leads to energy losses, possible overheating and rapid failure of the energy consumer.

The instantaneous power used by the consumer, respectively, is

$$P(t) = iu = \frac{U_0^2 \cos^2 \Omega t}{R} \qquad (5.136)$$

Average power emitted during the period of oscillation $P_{av}(t)$

$$P_{cp}(t) = \frac{1}{T}\int_0^T \frac{U_m^2}{2R}[1 + \cos 2\Omega t]dt = \frac{U_m^2}{2R} \qquad (5.137)$$

It is difficult to measure voltage and current amplitudes with electrical measuring instruments, because the instruments do not react to

the instantaneous value, but to the average value of the measured quantities. However, the average value of the harmonic-varying values of current and voltage over a period, is zero. Therefore, instead of measuring the mean value of the value itself, the mean value of its square (RMS value) is resorted to, namely

$(i^2)_{av}$ or $(u^2)_{av}$

The RMS value is independent of the current direction and therefore not zero. Calculation of the RMS values is carried out in the same way as the power P_{av}.

$$P_{av}(t) = \frac{1}{T}\int_0^T \frac{U_m^2}{2R}[1 + \cos 2\Omega t]dt = \frac{U_m^2}{2R} \tag{5.137}$$

$$(i^2)_{av} = \frac{I_m^2}{2};\ (u^2)_{av} = \frac{U^2}{2} \tag{5.138}$$

Quantities equal to $\sqrt{(i^2)_{av}}$ or $\sqrt{(u^2)_{av}}$ are called **effective values** of alternating current or voltage, and are denoted by *I or U*, respectively.

It follows from (5.138) that

$$I = \frac{I_m}{\sqrt{2}}; U = \frac{U_m}{\sqrt{2}} \tag{5.139}$$

Substituting U_m from (5.139) into (5.137), we obtain

$$P_{av} = \frac{2U^2}{2R} = \frac{U^2}{R}$$

as well as

$$P_{av} = IU \tag{5.140}$$

Comparing (5.140) with direct current power, we conclude that the effective current or voltage of an alternating current with an active resistance is respectively equal to the current and voltage of a direct current causing the same action as an alternating current.

5.2.2. Calculation of a circuit with capacitance

An alternating current circuit with capacitance is shown in Figure 5.16. The capacitance included in series in an electric circuit breaks it and therefore there is no direct current in it.

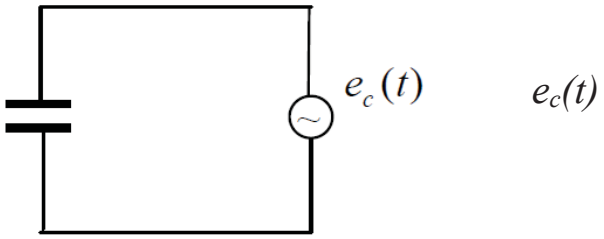


Figure 5.16.

In an alternating current circuit, the capacitance is continuously recharged due to changes in current direction, so the instantaneous and effective current values are not zero, although the average current per period is zero. Let's calculate the instantaneous current. According to Ohm's law for a circuit section

$$\hat{\imath}(t) = \frac{\hat{u}(t)}{\hat{X}_C} = \frac{U_m}{-j\dfrac{1}{\Omega C}}$$

$$\hat{\imath}(t) = jU_m\Omega C = U_m\Omega C e^{j\frac{\pi}{2}}$$

$$i(t) = I_{mc}\cos\left(\omega t + \frac{\pi}{2}\right) \qquad (5.141)$$

or

$$i(t) = -I_{mc}\sin\omega t \qquad (5.142)$$

where

$$I_{mc} = U_m\Omega C = \frac{U_m}{\hat{X}_c} \qquad (5.143)$$

From equation (5.141) it follows that the current on the capacitor is ahead of the voltage by an angle π / 2 (a quarter of a period).

5.2.3. Calculation of a circuit with inductance

An alternating current circuit with inductance is shown in Figure 5.17.

Figure 5.17.

Similar to the previous one

$$\hat{\imath}(t) = \frac{U_m}{\hat{X}_c} = \frac{U_m}{j\Omega L}$$

$$\hat{\imath}(t) = \frac{U_m}{\Omega L} e^{-j\frac{\pi}{2}} = I_m e^{-j\frac{\pi}{2}}$$

$$i(t) = I_m \cos\left(\omega t - \frac{\pi}{2}\right) = I_m sin\omega t$$

In a circuit with inductance, the inductive current lags the voltage in phase by an angle of $\pi / 2$ (quarter period).

5.2.4. Calculation of the complete RLC circuit

In a real RLC circuit, the current

$$\hat{\imath}(t) = \frac{\hat{u}(t)}{\hat{Z}}$$

In this case the impedance $\hat{Z}$ leads to a phase shift of the current with respect to the voltage by an angle f. This angle is determined, as can be seen from (5.131), by all the resistances included in the impedance $\hat{Z}$. In complex form

$$\hat{I} = \frac{\hat{U}}{\hat{Z}}$$

This equation, like the expression

$$\hat{\imath}(t) = \frac{\hat{U}(t)}{\hat{Z}}$$

is called **Ohm's law for an alternating current circuit** in complex form.

The most common methods of calculating alternating current electrical circuits are the vector diagram method and the complex method. The essence of the vector diagrams method has been described above, so let's consider the complex method of calculation.

Kirchhoff's laws, including the method of mesh currents, are used to calculate alternating current circuits, as in the case of direct currents. The calculation equations are made in the usual way with the only difference that a complex representation of values is used. In this case voltages, emf and currents are written in the equations after reduction by the multiplier eJ0}t. To avoid unnecessary calculations, only the initial phases of currents, voltages, and emf are entered into the complex exponents. The inductive and capacitive resistances are also written in complex form. The voltage drops are calculated by multiplying the corresponding currents by the resistances.

Using the conventional algebraic methods of solving systems of equations, the complex mesh currents and, respectively, the complex circuit element currents are calculated. The results are multiplied by the multiplier $e^{j\omega t}$, the real and imaginary values are distinguished, and their amplitudes and phases are calculated.

Examples of calculating alternating current circuits are given in the section on solving the corresponding problems.

5.3. Application of alternating current

It is known that of all types of energy, alternating current electricity is the most widespread in technology, industry and transport, as well as at home. This is due to the convenience of use and ease of delivery of this energy to the consumer.

The long-known direct current is not widely used, since no sufficiently powerful, efficient direct current sources have been built to date. Electric batteries and direct-current batteries have low capacity and power, are bulky and need frequent recharging, and deteriorate quickly. Direct current of relatively high power cannot be transmitted directly over long distances because of thermal losses in power lines.

The use of alternating current began at the beginning of the last century, when the corresponding electromechanical induction generator was invented. In principle, it is a frame rotating in a magnetic field. The electromagnetic induction emf arising in the frame creates electromagnetic oscillations in electrical circuits, i.e. alternating conduction current. The current changes in a sinusoidal, i.e. harmonic law with a frequency that is determined by the rotation speed of the frame. The value of the generated electricity and the generator power is determined by the magnetic field. The magnetic field is created by both permanent magnets (in low power generators) and electromagnets. The magnitude of the induction in the windings supplying the electric circuits of the consumers is determined by the number of turns of these windings.

The electromagnetic generator, included in the circuit, creates in it a powerful undamped forced oscillations. However, it is not possible to obtain current of high enough frequency, because it would require very high rotor speed of the generator. And in practice, high frequencies are not used in electrical engineering. The frequency of alternating current used in machinery and households is called the **industrial frequency**. This frequency within each country is standard, making it possible to create a unified energy grid. For example, in Russia and Israel the industrial frequency is 50 Hz, in some Western countries it ranges from 45 to 60 Hz.

Several types of induction generators are currently used to generate alternating current electricity. They all consist of an electromagnet or permanent magnet, a magnetic field source, and a winding in which alternating emf is induced. Structurally, an alternator consists of a stator and a rotor. In low-power alternators, the winding in which the emf is induced is placed in the rotor slots and the electromagnet

winding is placed in the stator slotsIn high power alternators, on the contrary, the solenoid winding is placed in the rotor, because it is more convenient to draw the generated current from the fixed winding. To create a magnetic field in the windings of the electromagnet, a relatively small current produced by an autonomous DC generator, the rotor of which is mounted on the shaft of the main generator, is fed with the help of so-called graphite brushes.

Industrial generators are installed at power plants. Power plants can be thermal, hydraulic or nuclear. In thermal power plants, the rotor of the generator is rotated by a steam turbine (turbine generator) or a heat engine. In a hydroelectric power plant, the generator rotor is rotated by a hydro turbine (hydroelectric generator). In nuclear power plants, a steam turbine powered by the heat of the nuclear reactor is used to drive the generator rotor.

Recently, a lot of attention has been paid to solar and wind power plants. In the future, it is expected to build fusion power plants that will use cheap and highly environmentally friendly fusion energy.

5.4. Power transmission over a distance

At the beginning of the 20th century, when the widespread use of electricity began, many small power plants were built for a particular consumer or small group of consumers. Transmission of electricity from the power plant was done over short distances and did not cause any major problems. Over time, however, it has become clear that it is much more profitable to build powerful power plants connected to the unified power grid of large industrial districts or even the entire country. The main problems in this regard were the losses of electricity during transmission over long distances.

These are thermal losses on the resistance of wires, losses associated with magnetic hysteresis, spin effect, gas discharges around overhead transmission lines, the influence of the reactive component of the resistance of transmission lines, reducing $cos\varphi$, useful power and a number of others. Thermal losses increase when transmitting large

volumes of low voltage currents (proportional to the square of the current strength), so it is much more advantageous to use high-voltage transmission lines with relatively low current values.

Transformers are used to convert current of a given wattage by increasing the voltage with a corresponding decrease in the amperage. The transformer was invented at the beginning of the 20th century. In principle, the transformer consists of a ferromagnetic core on which there are two windings with a small and a large number of turns. Due to the phenomenon of electromagnetic induction, the low voltage that is applied to the small winding is converted to a high voltage that is removed from the large winding, according to the ratio

$$\frac{U_1}{U_2} = \frac{N_1}{N_2} = K$$

where K is the transformation ratio;

N_1, N_2 - number of winding turns;

U_1, U_2 - voltages on the windings.

If a step-up transformer is used at the beginning of the line, a step-down transformer is used at the end of the line to supply consumers with low and therefore relatively safe voltage, e.g. 220 V.

The invention of three-phase systems further improved the profitability and efficiency of electricity transmission and use. A generator with an electromagnet consisting of three windings arranged at an angle of 120° with a corresponding phase shift of their emf is used there. Thus a rotating magnetic field is generated in the stator where the windings are located.

It is also possible to reduce losses by transmitting direct current instead of alternating current. However, since direct current cannot be transformed as alternating current, in this case you must use a system with special converters.

The power plant generates three-phase low-voltage current and transforms it into high-voltage current. The current is then converted by powerful mercury-arc rectifiers on ignitrons into high voltage direct

current and fed into the high voltage transmission line. On the receiving side, the process is reversed - the conversion of DC into AC with a voltage that is safe for the consumer.

Another highly efficient way to reduce losses while reducing the number of high-voltage overhead line networks is to replace them with cable lines that are laid in the ground. The widespread use of gas-filled cables, gas-insulated cables, superconducting and cryoconducting power cables significantly reduces power losses during transmission over long distances and increases the efficiency of the power transmission system.

5.5. Typical problems on oscillations

5.5.1. Problems on mechanical oscillations

Problem 1. A mathematical pendulum of length l = 1m. oscillates with amplitude A = 1 cm. In what time t_1 will the pendulum travel a path equal to its amplitude, if at the initial moment it was in a state of equilibrium? In what time t_2 and t_3 will the pendulum pass the first and second half of this path?

Solution. Since the amplitude of oscillations A<< l, the oscillations are harmonic and obey the equation

$$x = A \sin \omega_0 t$$

where A is the amplitude;

ω_0 is the natural cyclic frequency of oscillations.

$$\omega_0 = \sqrt{\frac{g}{l}}$$

If x = A , then

$$\sin \omega_0 t_1 = 1;\ \omega_0 t_1 = \frac{\pi}{2};\ \frac{2\pi}{2} t_1 = \frac{\pi}{2};\ t_1 = \frac{T}{4}$$

On the other side,

$$T = 2\pi\sqrt{\frac{1}{g}} = 2\pi\sqrt{\frac{1}{9{,}81}} \approx 2\ sec$$

Therefore

$$t_1 = 0{,}5\ sec$$

If $x = \frac{A}{2}$, то

$$\frac{A}{2} = A\sin\omega_0 t_2\,, \sin\omega_0 t_2 = \frac{1}{2}, \omega_0 t_2 = \frac{\pi}{6}$$

Therefore

$$t_2 = \frac{\pi T}{6 \cdot 2\pi} \approx 0.17\ sec$$

The second half of the path the pendulum will pass in time

$$t_3 = \left(\frac{T}{4} - 0{,}17\right) = 0{,}33\ sec$$

Problem 2. The sledge of length *l* slides through the snow without friction, but stops on the asphalt, whose coefficient of friction is μ. The weight of the sled is evenly distributed along its length. What is the braking time of the sled?

Solution. The equation of motion in this case is

$$lm_0\ddot{x} = -F_{TP}$$

where m_0 is the mass of a unit of sled length;

F_{fr} is the friction force of the sled on the asphalt.

The friction force by definition is equal to the product of the normal pressure force by the friction coefficient. Let us denote by x the part of the length of the sledge hit the asphalt equal to the value of the braking distance. The force of normal pressure is equal to the force of gravity F_g of the part of the sledge driven on the asphalt. Since

$$F_t = m_0 g x,$$

then

$$lm_0\ddot{x} = -\mu m_0 x g$$

or

$$\ddot{x} = \frac{\mu g}{l} x = 0$$

The resulting equality is the equation of harmonic oscillations, for which the natural frequency

$$\omega_0 = \sqrt{\frac{\mu g}{l}}$$

and period

$$T = 2\pi \sqrt{\frac{l}{\mu g}}$$

Before stopping the sledge passes a quarter of the period of oscillation, therefore

$$t = \frac{T}{4} = \frac{\pi}{2} \sqrt{\frac{l}{\mu g}}$$

Problem 3. A variable force with a frequency of 16 Hz is applied to the end of the spring of a spring pendulum with a mass m = 1kg. Will there be resonance if the spring stiffness k = 400 N/m.

Solution. Resonance occurs when the frequency Ω of the forcing force coincides with the natural frequency ω_0. The value of ω_0 for the spring is

$$\omega_0 = \sqrt{\frac{k}{m}} = \sqrt{\frac{400}{1}} = 20\ Hz$$

Thus, $\Omega < \omega_0$ and therefore no resonance occurs in this system.

Problem 4. Frequency of the natural vertical oscillations of the railroad car ω_0. On the joints of the rails, the car receives periodic repetitive shocks. At what speed of the train resonance can occur if the length of each rail between joints l.

Solution. According to the condition, the necessary speed of the train

$$\upsilon = \frac{l}{T}$$

where T is the period of natural oscillations of the car;

ω_0 is the frequency of natural oscillations.

Since

$$T = \frac{2\pi}{\omega_0}$$

then substitution gives

$$\nu = \frac{l\omega_0}{2\pi}$$

5.5.2. Problems on oscillations in an oscillating circuit

Problem 5. What are the cyclic natural frequency of oscillations, frequency of damped oscillations, damping coefficient, logarithmic decrement of damping, initial phase of oscillations of the series circuit RLC with R = 50 Ohm, L = 100 μH, C = 5 nF. Determine the mode of oscillations in the circuit

Solution. The natural frequency of the oscillating circuit

$$\omega_0 = \sqrt{\frac{1}{LC}}$$

Substitution gives ω_0 = 1.43 MHz.

Frequency of oscillations in the circuit, taking into account dissipations (damped oscillations)

$$\omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} = 1{,}32\ MHz$$

Attenuation coefficient δ and logarithmic decrement of damping χ

$$\delta = \frac{R}{2L} = 0{,}5 \cdot 10^6\ sec^{-1};\ \chi = \delta T = 0{,}5 \cdot 10^6 \frac{2\pi}{1.32 \cdot 10^6} = 2{,}42$$

The oscillation mode is determined by the ratio between the active and reactive resistance of the circuit. If $R < \frac{1}{2}\sqrt{X_L X_C}$, then the mode in the circuit is oscillatory. If $R > \frac{1}{2}\sqrt{X_L X_C}$, the mode is aperiodic.

$$\sqrt{X_L X_C} = \sqrt{\frac{10^{-4}}{5 \cdot 10^{-9}}} = \sqrt{0{,}2 \cdot 10^5} = 140\ \Omega$$

Since $R < \frac{1}{2}\sqrt{X_L X_C}$ the mode in the circuit is oscillatory.

The initial phase of the oscillations

$$\omega_0 = \tan^{-1}\sqrt{\frac{4L}{R^2} = -1} = \tan^{-1}\sqrt{\frac{4L \cdot 10^{-4}}{10^4 \cdot 10^{-9}} - 1} = \tan^{-1} 6{,}38 \cong 81° \cong 0{,}9\frac{\pi}{2}$$

5.5.3. Problems on alternating current

Problem 6. An alternating current is flowing in the RLC circuit

$$i = I_m cos\omega t$$

Calculate the amplitude values of the voltage drop across each of the circuit elements and the instantaneous value of the voltage across the circuit terminals, and determine the law of variation of this voltage.

Solution. According to Ohm's law for a circuit section, the voltage drop in that section is equal to the product of the current by its resistance, so

$$U_{mR} = I_m R;\ U_{mX_L} = I_m \omega L;\ U_{mX_c} = I_m \frac{1}{\omega C},$$

where the *"m"* sign denotes the amplitude value of the parameter $U_m=I_mZ$,

Z is the circuit resistance

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2};$$

Um is the voltage drop across the terminals of the complete RLC circuit.

Since the instantaneous current value

$$i = I_m \cos \omega t,$$

then in the complex form

$$\hat{\imath} = I_m$$

The complex resistance of the circuit and the voltage drop are determined from the formulas

$$\hat{Z} = R + j\left(\omega L - \frac{1}{\omega C}\right) = Z_m e^{j\omega}$$

where ω *and* Z_m are determined from the equations

$$tg\varphi = \frac{\omega L - \frac{1}{\omega C}}{R};$$

$$Z_m = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2};$$

a

$$\hat{u} = \hat{\imath}\hat{Z} = I_m Z_m e^{j\varphi} = U_m e^{j\varphi}$$

$$U = Re\hat{u} = U_m \cos(\omega t + \varphi)$$

where

$$\varphi = \tan^{-1} \frac{\omega L - \frac{1}{\omega C}}{R};$$

$$U_m = I_m Z_m$$

Problem 7. The electric circuit of the oscillating circuit includes in series: a resistor of 1 kOhm, a capacitor of 1 μF and a coil with an inductance of 0.5 H. Find the resonant frequency ν_p, reactances X_L *and* X_C, impedance Z at the frequency of the emf source ω = 62.8 kHz.

Solution.

$$X_C = \frac{1}{\omega C} = \frac{1}{62{,}8 \cdot 10^3 \cdot 10^{-6}} = 16\ \Omega$$

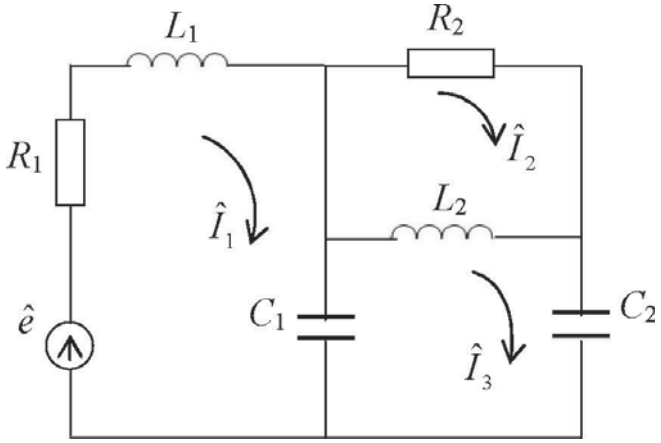
$$X_L = \omega L = 62{,}8 \cdot 10^3 \cdot 0{,}5 = 31{,}4\ \Omega$$

$$\mathrm{Z} = \sqrt{10^6 + (31{,}4 \cdot 10^3 - 16)^2} \cong 31{,}4\ \Omega$$

$$\nu_\rho = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{6{,}28\sqrt{10^{-6} \cdot 0{,}5}} = 225\ Hz$$

Problem 8. Calculate the AC circuit of industrial frequency ν = *50 Hz*, shown in the figure, if $e = 2\sin\omega t$;

$R_1 = R_2 = 5\ Ohm;\ L_1 = 5\ mH;\ C_1 = 1\ mF\ ;\ L_2 = 10\ mH;\ C_2 = 0{,}5\ mF;$



Solution. Let us apply the method of mesh currents. The equations for mesh currents in complex form are

$$\hat{I}_1\left(R_1 + j\omega L_1 - j\frac{1}{\omega L_1}\right) - \hat{I}_2 \cdot 0 + \hat{I}_3 j\ \frac{1}{\omega C_1} = \hat{e}$$

$$-\hat{I}_1 \cdot 0 + \ \hat{I}_2(R_2 + j\omega L_2) - \hat{I}_3 j\ \omega L_2 = 0$$

$$+\hat{I}_1 j\frac{1}{\omega C_1} - \hat{I}_2 j\,\omega L_2 + \hat{I}_3\left(j\,\omega L_2 - j\frac{1}{\omega C_2} - j\frac{1}{\omega C_1}\right) = 0$$

After substituting the numerical values, we obtain

$$(5-1.6j)\hat{I}_1 - 0\cdot\hat{I}_2 + 3{,}18j\hat{I}_3 = 2$$

$$0\cdot\hat{I}_1 + (5+3{,}14j)\hat{I}_2 - 3{,}14j\hat{I}_3 = 0.$$

$$3.18j\cdot\hat{I}_1 - 3.14j\hat{I}_2 - 6{,}4j\hat{I}_3 = 0$$

We solve the system by the determinant method

$$\hat{I}_1 = \frac{\Delta_1}{\Delta};\ \hat{I}_2 = \frac{\Delta_2}{\Delta};\ \hat{I}_3 = \frac{\Delta_3}{\Delta}$$

System determinant

$$\Delta = \begin{vmatrix} (5-1{,}6j) & 0 & 3{,}18j \\ 0 & (5+3{,}14j) & -3{,}14j \\ 3{,}18j & -3{,}14j & -6{,}4j \end{vmatrix} \cong (1{,}5-1{,}75j)10^2$$

$$\Delta_1 = \begin{vmatrix} 2 & 0 & 3{,}18j \\ 0 & (5+3{,}14j) & -3{,}14j \\ 0 & -3{,}14j & -6{,}4j \end{vmatrix} = 60-64j$$

$$\Delta_2 = \begin{vmatrix} (5-1{,}6j) & 2 & 3{,}18j \\ 0 & 0 & -3{,}14j \\ 3{,}18j & 0 & -6{,}4j \end{vmatrix} = 20$$

$$\Delta_3 = \begin{vmatrix} (5-1{,}6j) & 0 & 2 \\ 0 & (5+3{,}14j) & 0 \\ 3{,}18j & -3{,}14j & 0 \end{vmatrix} = 20-32j$$

$$\hat{I}_1 = \frac{60-64j}{150-175j} = \frac{(60-64j)(150+175j)}{150^2+175^2} = \frac{2+0{,}09j}{5{,}3}$$

$$\cong 0{,}38^{0{,}015j}$$

$$\hat{I}_2 = \frac{0{,}3 + 0{,}35j}{5{,}3} \cong 0{,}09e^{0{,}27\pi j}$$

$$\hat{I}_3 = \frac{0{,}86 + 0{,}13j}{5{,}3} \cong 0{,}16e^{-0{,}05\pi j}$$

$$i_1 = i_{R1} = i_{L1} = 0{,}38\cos(\omega t + 0{,}015\pi)$$

$$i_2 = i_{R2} = 0{,}09\cos(\omega t + 0{,}027\pi)$$

$$i_3 = i_{C2} = 0{,}16\cos(\omega t + 0{,}05\pi)$$

$$\hat{I}_{C1} = \hat{I}_1 - \hat{I}_3 = 0{,}38 + 0{,}017j - 0{,}16 + 0{,}025j = 0{,}22 + 0{,}042j$$

$$\hat{I}_{C1} = 0{,}22\cos(\omega t + 0{,}06\pi)$$

$$\hat{I}_{L1} = \hat{I}_2 - \hat{I}_3 = 0{,}056 + 0{,}06j - 0{,}16 + 0{,}025j = -0{,}1 + 0{,}03j$$

$$\hat{I}_{L2} \cong -0{,}1\cos(\omega t - 0{,}05\pi)$$

Problem 9. Determine the dependence of the specific electrical conductivity of intrinsic semiconductors on temperature.

Solution. Specific conductivity γ can be determined from Ohm's law, according to which the dependence of the current density vector $\vec{J}$ on the electric field strength $\vec{E}$ is given by the equation

$$\vec{J} = \gamma\vec{E}.$$

On the other hand, since the current in the intrinsic semiconductor is the sum of electron and hole currents with densities $\vec{J}_n$ and $\vec{J}_p$, which

are created by the motion of carriers with average velocities υ_{nav} *and* υ_{pav}, and effective masses m_n^* m_p^* then

$$\vec{J} = n_n(-e)\vec{v}_{nav} + n_p(+e)\vec{v}_{pav}$$

and

$$\vec{J} = e^2\left(\frac{n_n\tau_n}{m_n^*} + \frac{n_p\tau_p}{m_p^*}\right)\vec{E}$$

Since holes and electrons in the intrinsic semiconductor generate and recombine simultaneously with equal probability, their concentrations are equal and proportional to this probability. Thus, comparing this expression with Ohm's law and considering that $n_n = n_p = n$ and denoting

$$e^2\left(\frac{\tau_n}{m_n^*} + \frac{\tau_p}{m_p^*}\right) = \gamma_0,$$

we obtain that

$$\gamma = \gamma_0 n$$

Generation is the result of an electron moving from the valence band to the conduction band by jumping over the band gap. According to Boltzmann's law, the probability of generation ρ is proportional to temperature T and inversely proportional to the band gap width ΔE_3

$$p \approx e^{\frac{\Delta E_3}{kT}}$$

Finally

$$\gamma \approx \gamma_0 e^{\frac{\Delta E_3}{kT}}$$

Chapter 6. Wave processes

The perturbation of physical quantities that propagate in space, which characterize the states of physical objects, transferring the energy and nature of the perturbation from one point of space to another, is called a **wave**. It is believed that waves propagate in a **continuous** medium although the idea of a continuous medium is an idealization of reality, since purely continuous media do not exist in Nature. Waves can carry both **oscillatory** and **non-oscillatory** motion. According to this definition, waves that carry oscillatory motion can ideally be considered as collective oscillations of a continuous medium filled with an infinite number of interconnected **point oscillators**. Due to the fact that the description of a continuous medium requires specifying its defining parameters in an infinite number of points in space, such a medium is characterized by an infinite number of degrees of freedom.

When considering a propagating wave, it is always possible to distinguish a set of points in the medium whose changes are in the same phase at a given moment. The totality of these points forms an **constant phase surface** in space, which is also called a **wave surface**. When this surface is in contact with an unexcited medium, it is called a **wave front**. The wave surface and the wave front at points of homogeneous and isotropic space have a spherical shape. Such waves are called **spherical waves**. Sources of spherical waves are called **point** sources. Waves whose constant phase surface is a plane are called **plane waves**. A line perpendicular to the wave surface (wave front) is called a **ray**. Rays of spherical waves are radial lines of the wave surface.

The rays of plane waves are beams of parallel lines perpendicular to the plane of the wave front. Obviously, the beams of wave rays coming from a distant source can always be assumed to be parallel and the waves to be flat. According to the definition, the rays on average coincide with the direction of wave propagation. If a wave propagates along the direction of oscillation, it is called a **longitudinal wave**. A wave that propagates perpendicular to the direction of oscillation is called a **transverse wave**. It is obvious that longitudinal waves by this definition

are, in particular, waves excited by compressive and tensile strains in a real medium. Waves excited by the shear deformation of the material medium are, on the contrary, transverse.

Examples of longitudinal waves in a material medium are sound waves, and in an immaterial medium, presumably, gravitational waves. Examples of transverse waves are waves propagating in a sounding string or in a stretched cord, as well as electromagnetic waves, etc.

The propagation of the wave is not related to the movement of particles directed along it. The wave process only transfers energy. The result is an energy flow that diverges in all directions from the source. This flow is called a **Poynting vector**. The direction of the Poynting vector coincides with the direction of wave propagation. The energy per unit of wave surface passing through an observation point per unit time is called the **intensity of the wave**. The intensity is calculated by the formula

$$J = \frac{N}{S} \qquad (6.1)$$

where

N is the power carried by the wave;

S is the area of the spherical wave surface.

Since $S = 4\pi R^2$, the intensity of the wave is always inversely proportional to the square of the distance from the source to the observation point.

The representation of continuous media is justified only in the classical approximation, which is limited to the consideration of relatively slow macro-processes, in the description of which the discrete structure of physical bodies and fields can be neglected. From the classical point of view, all processes occurring in nature and connected with energy exchange are reduced to the motion or deformation of discrete (corpuscular) material bodies, obeying the laws of mechanics, or - to the fundamentally different from them spread of wave processes in continuous media.

Meanwhile, by the end of the 19th and beginning of the 20th century, facts began to accumulate that increasingly demonstrated that there are neither completely discrete nor completely continuous structures in the real world.

In particular, it turned out that there is no fundamental difference between the real behavior of particles, which in classical physics were attributed purely corpuscular properties resulting from their discreteness, and the real propagation of processes in continuous media, which were attributed purely wave properties resulting from the continuity of the medium.

In particular, it turned out that real objects are neither pure corpuscles (particles) nor pure waves. They actually have both corpuscular and wave properties. This concept was called **wave-particle duality** and formed the basis of a new, quantum physical theory.

According to quantum theory, objects can generally exhibit both corpuscular and wave properties to a greater or lesser extent, depending on the conditions of the processes in which they are involved. Therefore, before proceeding to the study of real processes, it is necessary, following the already considered behavior of discrete bodies, to consider in detail purely wave processes. Let us limit ourselves to the consideration of waves that carry oscillatory processes.

6.1. Classification of waves

The material medium is known to consist of particles bound together by electromagnetic interactions. When such a medium is mechanically perturbed, i.e. when any of its points is removed from the equilibrium state, various types of deformation arise in it, which are transmitted from one particle to another particle associated with it in the form of waves. Such waves are called **elastic waves**. Depending on the type of strain transmitted by the elastic wave, it may be longitudinal or transverse. Elastic waves, as mentioned above, transmitting compressive or tensile strain of the medium are longitudinal, and waves transmitting shear strain are transverse. Since expansion and contraction deformations

are always elastic (see Section 2.9), the propagation of elastic longitudinal waves can occur in any material medium, regardless of its aggregate state. On the contrary, shear strains in liquids, gases and plasma are known to be inelastic. For this reason, elastic transverse waves can propagate only in solids (see point 2.9). Elastic waves, by their definition, are **mechanical waves** because they are always transferred from one particle to other particles associated with it. This means that elastic waves can only propagate in a real medium.

In real medium, in addition to elastic waves, so-called capillary (**surface**) waves can also propagate. The latter occur, in particular, on the surface of the liquid. The source of such waves are external disturbances that remove the medium from the state of equilibrium. The forces that restore equilibrium to the medium are surface tension or gravity.

Wave processes are not necessarily associated with mechanical motions that perturb the medium. In some cases, there may be perturbations that result from changes in non-mechanical properties of objects, such as electrical, magnetic, optical, etc.

The waves arising in this case are of a purely non-mechanical nature. Examples of **non-mechanical** waves are electromagnetic waves, as well as hypothetical gravitational waves not yet detected by direct observation. Non-mechanical waves, according to their definition, can propagate in both real and immaterial medium, such as vacuum.

In quantum theory, which, as already mentioned, is based on the idea of wave-particle duality, elementary particles are compared with waves with special properties. These waves are called **matter waves**. Sometimes matter waves are referred to as de Broglie waves after Louis de Broglie, who first introduced them into consideration. On the contrary, light waves, as well as other types of radiation have corpuscular properties and are considered, under certain conditions, as particles called **photons**.

Electromagnetic waves were predicted by Faraday. Maxwell, in turn, justified theoretically the possibility of their existence (see section 4.2.3) in his famous electromagnetic field theory. In particular, he showed that the electromagnetic field is the source of fluctuations in the properties

of the medium, which are not necessarily related to the movement of its particles. Therefore, for the propagation of electromagnetic waves, the presence of a material medium is not required. Electromagnetic waves were experimentally discovered and studied by Hertz.

Maxwell also drew attention to the fact that the so-called electrodynamic constant c included in the equations of the electromagnetic field determines, on the one hand, is the speed of propagation of electromagnetic waves and, on the other hand, coincides with the speed of light in a vacuum, measured experimentally, with a very high accuracy. On this basis, Maxwell proposed the idea that light has the character of electromagnetic waves. Let us point out, in this connection, that the wave theory of light was put forward in the 17th century by Descartes. This theory was further developed by Huygens and Fresnel. Huygens believed that light waves, by analogy with sound, are longitudinal.

He, however, like many other scientists of his time, mistakenly believed that any waves could propagate only in an elastic, material medium. Since light, unlike sound, also propagated in the vacuum, he, like Descartes in his time, believed that the vacuum is filled with some elastic continuous (continuous) material medium, which was called the world (light-carrying) ether.

Conceptions of the world's ether initially encountered serious difficulties. First of all, it was necessary to explain why all bodies, including celestial bodies, move freely relative to the ether without encountering resistance from it. Up until Maxwell's theory, this phenomenon was explained by the fact that light, like sound, has the character of longitudinal waves, and the ether is, although elastic, but so rarefied medium that the resistance it offers to the movement of bodies is negligible and virtually imperceptible. However, Fresnel's experiments, later confirmed by Maxwell's theory, established the transverse nature of light waves.

In this regard, it was necessary to reconsider the above concepts of the ether, because otherwise it would be necessary to make the absurd assumption that the ether is an absolutely solid medium. In addition, the

notion of the ether as a rarefied medium was in no way consistent with the enormous speed of propagation of electromagnetic waves, including light. The result was a new, albeit poorly understood model, according to which the ether was represented as a continuous but inelastic medium. In this case, light was seen as surface waves, similar to the waves on the surface of a liquid. The surface tension of the liquid was likened to the tension of the electric and magnetic field lines of a light wave.

Another difficulty associated with the ether arose as the idea of the existence of absolute rest and motion re-entered physics along with it. The fact is that, according to the laws of electrodynamics established by Maxwell's theory, the speed of propagation of light in a vacuum is the same everywhere.

This meant that the ether that fills the vacuum is an absolutely stationary system of reference, and the speed of light in the vacuum is the absolute speed. It followed that the principle of relativity, established in mechanics by Galileo, is not fulfilled in electrodynamics.

In other words, the speed of light emitted by a source moving relative to the stationary ether must have changed with the change in speed and direction of the light source.

At the end of the 19th century, physicists Albert A. Michelson and Edward W. Morley set up an experiment by which the speed of light was measured in the direction of the Earth's motion and in the perpendicular direction. However, very careful and very accurate measurements of the speed of light, repeated later by other researchers, found no difference between these speeds.

Thus, the Michelson-Morley experiments disproved the possibility of the existence of an absolutely stationary world ether. The alternative assumption that the ether, as an all-permeating medium, and therefore light, are completely entrained by moving bodies, also had to be discarded, as it contradicted the experiment.

The way out of this seemingly hopeless situation was suggested by Einstein. The special theory of relativity developed by him in 1905 allowed, first, to abandon the absolute space and time introduced by

Newton, and, second, it led to the rejection of the existence of an absolutely stationary reference system associated with the ether. This, in turn, allowed Einstein to abolish the ether itself, now completely superfluous, both from the standpoint of electromagnetic theory and, in particular, from the standpoint of relativity (see also Chapter 3), and to do away with the problems associated with it.

6.2. Wave function

Oscillations, as was shown above, are described by an oscillatory periodic function of time, so the wave process that carries oscillations can be described by the same function, but depending on both time and space coordinates.

Let us first consider the wave function in the classical approximation. For this purpose, let us assume that in some point of space-time filled with a continuous medium, the oscillations of a point oscillator are excited. These oscillations propagate in the form of a wave, passing from one point of the medium to another. This means that the oscillations of each successive point of the medium with respect to the previous one occur with some time lag by the value of t_1, or, what is the same, with some phase shift. From the theory of oscillations considered in the previous chapter, taking into account the specified lag, the state of each point in space can be described by means of the function

$$\psi(r,t) = f[\omega(t - t_1)] \qquad (6.2)$$

where ω is the angular frequency of oscillations of the source oscillator.

Since the function $\psi(r,t)$ reflects the character of oscillations in an arbitrary point of space, we can assume that it describes a wave process propagating in the medium. That is why it is called a **wave function**. If we assume that the medium is homogeneous and isotropic, then it is logical to assume that the wave propagates in space-time at a constant speed v. For simplicity, let us assume that the wave propagates in one-dimensional space in the direction of the Ox axis, and the source oscillator is located at the origin, then

$$t_1 = \frac{x}{\upsilon} \tag{6.3}$$

Substituting (6.3) into (6.2), we obtain that

$$\psi(x,t) = f\left[\omega\left(t - \frac{x}{\upsilon}\right)\right] \tag{6.4}$$

or

$$\psi(x,t) = f\left[\omega t - \frac{\omega}{\upsilon}x\right] \tag{6.5}$$

The value $\frac{\omega}{\upsilon}$ can be calculated in the following way

$$\frac{\omega}{\upsilon} = \frac{2\pi\nu}{\upsilon} = \frac{2\pi}{\upsilon T} \tag{6.6}$$

where

v is the linear frequency of oscillation of the oscillator;

T is the period of oscillation.

As the wave propagates in space-time, many points (oscillators) appear in which the lag time is a multiple of the oscillation period, i.e.

$$t_l = mT,$$

where *m = 1, 2, 3, ...*

For these points

$$x_{1m} = m\upsilon T \tag{6.7}$$

If we assume *m = 1*, then

$$x_{11} = \upsilon T \tag{6.8}$$

Thus, x_{11} is the distance between the two closest points, the time lag of oscillations in which is equal to the period T. This distance is called the wavelength and is denoted by λ.

It follows from (6.8) and (6.6) that

$$\lambda = \upsilon T \tag{6.9}$$

and

$$\frac{\omega}{\upsilon} = \frac{2\pi}{\lambda} = k \tag{6.10}$$

The value of k is called the **wave number**, and a vector equal in absolute value to the wave number and directed in the direction of wave propagation is called the wave vector $\vec{k}$.

Since, according to (6.10)

$$\frac{\omega}{k} = \upsilon$$

then

$$\psi(x,t) = f(\omega t - kx) \tag{6.11}$$

$$\psi(x,t) = f[k(\upsilon t - x)] \tag{6.12}$$

It follows from relation (6.11) that the frequency ω and the wave number k characterize the periodicity of the wave process in time and space, respectively.

The wave arising from the oscillations of a point oscillator propagates, generally speaking, in three-dimensional space. Therefore, in the general case

$$\psi(r,t) = f[\omega t - kr] \tag{6.13}$$

or

$$\psi(r,t) = f[k(\upsilon t - r)] \tag{6.14}$$

6.3. Wave equation

The wave function introduced by relations (6.13) and (6.14), as follows from the theory of differential equations of mathematical physics, satisfies the equation

$$\frac{\delta^2\psi}{\delta x^2}+\frac{\delta^2\psi}{\delta y^2}+\frac{\delta^2\psi}{\delta z^2}-\frac{1}{v^2}\frac{\delta^2\psi}{dt^2}=0 \tag{6.15}$$

In the particular case of wave propagation in one-dimensional space, equation (6.15) takes the form

$$\frac{\delta^2\psi}{\delta x^2}-\frac{1}{v^2}\frac{\delta^2\psi}{dt^2}=0 \tag{6.16}$$

Equation (6.15) or partial equation (6.16) is called the **wave equation**.

The quantity

$$\Delta\psi=\left(\frac{\delta^2}{\delta x^2}+\frac{\delta^2}{\delta y^2}+\frac{\delta^2}{\delta z^2}\right)\psi=\frac{\delta^2\psi}{\delta x^2}+\frac{\delta^2\psi}{\delta y^2}+\frac{\delta^2\psi}{\delta z^2}$$

is called the **Laplace operator** or Laplacian, and the value

$$\Delta=\frac{\delta^2}{\delta x^2}+\frac{\delta^2}{\delta y^2}+\frac{\delta^2}{\delta z^2}$$

is the Laplace differential operator (see Appendix 3). With this in mind, equation (6.15) is usually written in the form

$$\Delta\psi-\frac{1}{v^2}\frac{\delta^2\psi}{\delta t^2}=0 \tag{6.17}$$

The wave equation (6.16), in addition to the solution (6.12), also allows a solution

$$\psi(x,t)=f[k(vt+x)] \tag{6.18}$$

The physical meaning of this solution follows from the fact that for it

$$v=-\frac{dx}{dt} \tag{6.19}$$

This means that (6.18) can be seen as a function of the wave propagating in the direction opposite to the wave propagation expressed

by the relation (6.14). In this sense, a wave according to (6.12) is considered a **direct wave** and is called an incident wave, and a wave according to (6.18) is called a **reflected wave**.

The specific form of the wave function ψ, given by the function *f*, determines the form (profile) of the wave and depends on the law of oscillations of its source at the starting point $x = x_0 = 0$.

Among the various laws of oscillation, the most common are harmonic oscillations, which occur with **small** amplitude, and the function *f* is given by **cosine or sine** to the first power. Let us assume that

$$f(\omega t - kx)_{x=0} = A\cos\omega t \tag{6.20}$$

where

A is the amplitude of oscillations,

ω is the cyclic frequency of oscillations.

The following wave functions satisfy condition (6.20)

$$\left.\begin{aligned}\psi(x,t) &= \text{acos}\,(\omega t \pm kx)\\ \psi(x,t) &= \text{bsin}\,(\omega t \pm kx)\end{aligned}\right\} \tag{6.21}$$

Equation (6.21) can also be written as

$$\left.\begin{aligned}\psi(x,t) &= a\ \cos\left[2\pi\left(\frac{t}{T}-\frac{x}{\lambda}\right)\right]\\ \psi(r,t) &= b\cos\left[2\pi\left(\frac{t}{T}-\frac{r}{\lambda}\right)\right]\end{aligned}\right\} \tag{6.22}$$

or as any linear combination of these functions.

The waves described by equations (6.21) and (6.22) are called **harmonic waves**. They express an infinite periodic process in time and space (see Appendix 3). Harmonic waves are also called **sine waves** or **linear waves**.

The wave function $\psi(x,t))$ or $\psi(r,t)$ (see Appendix 3) can also be given a complex form as follows

$$\left.\begin{aligned}\psi(x,t) &= A[\cos(\omega t - kx) + j\sin(\omega t - kx)]\\ \psi(r,t) &= A[\cos(\omega t - kx) + j\sin(\omega t - kr)]\end{aligned}\right\} \quad (6.23)$$

Or, using the Euler formula (see Appendix 3), we can obtain the wave function from (6.23) in the complex form as

$$\left.\begin{aligned}\psi(x,t) &= Ae^{j(\omega t - kx)}\\ \psi(r,t) &= Be^{j(\omega t - kr)}\end{aligned}\right\} \quad (6.24)$$

6.4. Phase and group velocities. Wave packets

Let us call a single harmonic wave of a certain length λ **monochromatic**. Real waves are not monochromatic, but consist of a set (spectrum) of waves whose lengths lie between λ_2 and λ_1. The value $\Delta\lambda = \lambda_1 - \lambda_2$, *or* $\Delta k = k_1 - k_2$, is called the width of the spectrum. Real waves are described by decomposing them into a Fourier series.

For a wave process using the Euler formula, this decomposition takes the form

$$\psi(x,t)|_{t=const} = \sum_{n=-\infty}^{n=+\infty} a_n e^{j(\omega_n t - nkx)} \quad (6.25)$$

Let us denote

$$a_n e^{j\omega_n t}|_{t=const} = A_n$$

then

$$\psi(x,t)|_{t=const} = \sum_{n=-\infty}^{n=+\infty} A_n e^{-jnkx} \quad (6.26)$$

where according to Fourier series theory

$$A_n = \frac{1}{\lambda_n}\int_0^{\lambda_n} \psi(x,t)|_{t=const} \cdot e^{jnkx}\,dx \quad (6.27)$$

The decomposition (6.26) is called a spectral decomposition. The same decomposition can be performed over time when *x* = *const*.

Spectral decompositions of the wave function of an arbitrary wave process can be viewed as the result of addition (superposition) of harmonic waves with multiples of wave numbers or multiples of frequencies and different amplitudes. It follows from the above that any real wave process is nothing but a superposition of an infinite number of harmonic waves of multiple frequency or multiple wave number in the general case. If there is only one monochromatic wave as part of the spectrum then the width of the spectrum $\Delta k \to 0 (\Delta\omega \to 0)$ and the intervals of time and space in which the wave is defined tend to infinity, i.e. $\Delta x \to \infty,\ \Delta \mathrm{t} \to \infty$ (Fig.6.1a).

On the contrary, if the width of the spectrum tends to infinity, i.e. it includes an infinite number of monochromatic waves $\Delta k \to \infty, \text{или}\ \Delta\omega \to \infty$ then as follows from Fourier series theory the process is strictly localized in space or time. i.e. $\Delta x \to 0, \text{или}\ \Delta t \to 0$ (Figure 6.1b).

For wave packets

$$\left.\begin{aligned} \Delta k \cdot \Delta x &= a \\ \Delta\omega \cdot \Delta t &= b \end{aligned}\right\} \tag{6.28}$$

where a and b are constants. Since a ≥ 1 and b ≥ 1, (6.28) has the form

$$\left.\begin{aligned} \Delta k \cdot \Delta x &\geq 1 \\ \Delta\omega \cdot \Delta t &\geq 1 \end{aligned}\right\} \tag{6.29}$$

Of particular interest are the packets formed by a finite group of monochromatic waves close in wave numbers and frequencies, namely

$$\left.\begin{aligned} k_0 - \Delta k \leq k \leq k_0 + \Delta k \\ \omega_0 - \Delta\omega \leq \omega \leq \omega_0 + \Delta\omega \end{aligned}\right\} \tag{6.30}$$

where

$$\left.\begin{aligned} \Delta k &\leq k_0 \\ \Delta\omega &\leq \omega_0 \end{aligned}\right\} \tag{6.31}$$

Such packets are usually called a **wave train**. They can be defined as a propagating nonlinear wave occupying a limited region of space-time at any given time.

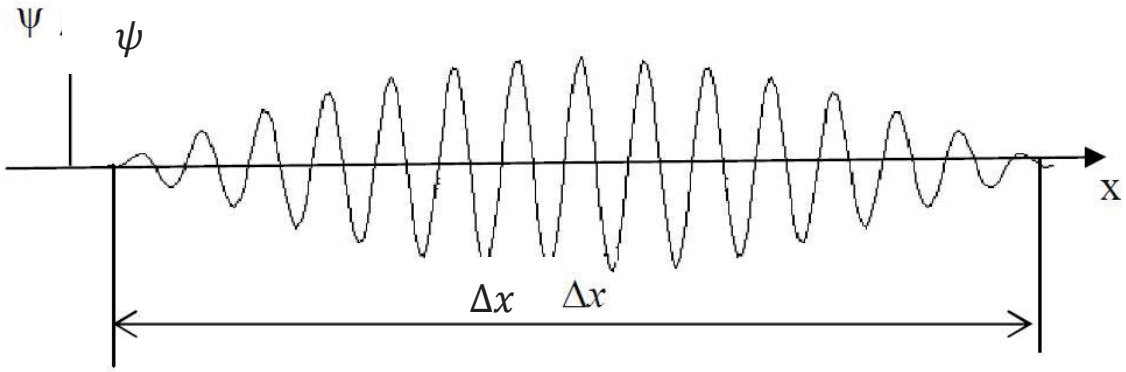


Figure 6.1a

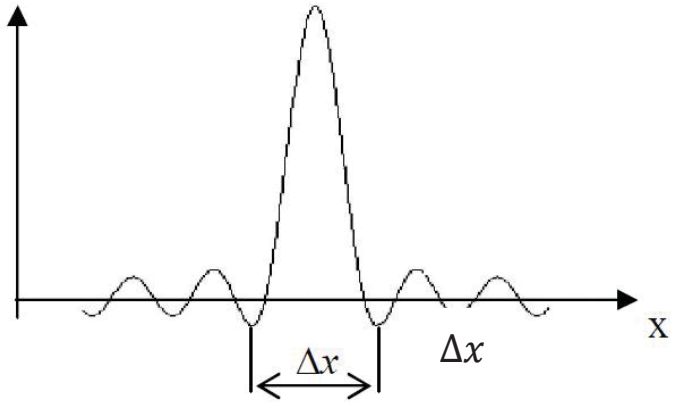


Figure 6.1b.

The wave train usually contains a limited number of wave crests (maxima) fairly well localized in space-time. In the limit, when the number of harmonics forming the wave train tends to infinity and the spectral composition of the wave train becomes continuous, there remains only one maximum, which is localized in one point of space and time. Mathematically, such a spectrum of harmonic waves describes a perfect **solitary wave**. For a long time, the solitary wave was considered an idealization of real reality. However, in 1834 John Scott Russell first observed such a wave on the surface of the water. It turned out that solitary waves can arise and exist really in the form of waves of different nature, but only under certain conditions and when they propagate in an inhomogeneous medium, which has, as will be shown below, quite certain properties. Real solitary waves, unlike ideal waves, are localized,

not in one point, but in some area of space. These solitary waves are called **solitons**. Solitons are non-linear waves. Under the influence of the properties of the medium the soliton forms a structurally stable wave, which in many ways behaves like a particle of matter.

A monochromatic wave infinite in time and space is an idealized abstraction that does not really exist, but is part of really existing waves.

Each of them is characterized by its unchanging on average in time and space phase

$$\varphi = \omega t - kx = const \tag{6.32}$$

The phase velocity of a wave is the speed of travel of a given monochromatic wave packet with a given phase. Therefore, to determine the phase velocity we differentiate in time the relation (6.32), then

$$\omega - k\frac{dx}{dt} = 0 \tag{6.33}$$

whence, by the definition of the phase velocity

$$\upsilon_f = \frac{dx}{dt} = \frac{\omega}{k} \tag{6.34}$$

The phase velocity coincides, obviously, with the velocity of the moving front of the monochromatic wave. If a wave propagates in a certain direction, for example, in the direction of the Ox axis, then its monochromatic components can move in any direction ρ such that

$$\rho = \mathrm{x}\cos\varphi \tag{6.35}$$

Let, for example, a plane wave propagate along the Ox axis with a velocity υ_x (Fig. 6.2), and the component along Oρ

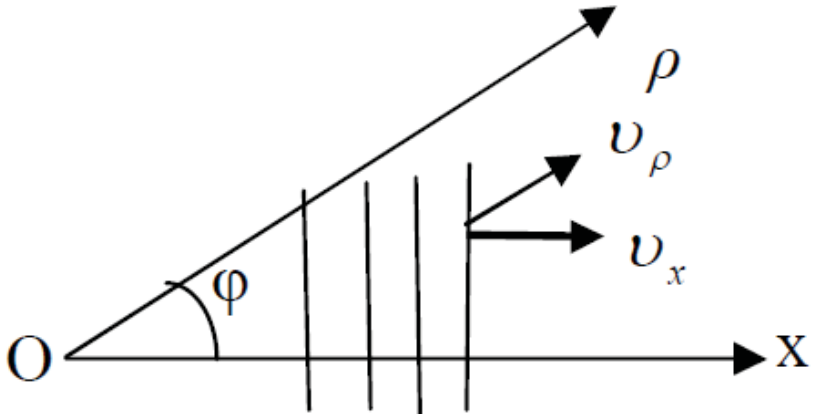


Figure 6.2.

Then the phase velocity

$$\upsilon_f = \frac{\upsilon_x}{\cos\varphi} \tag{6.36}$$

In other words, the phase velocity can generally exceed the wave propagation velocity υ_x.

If, say, the wave in question is light from a distant source that propagates with speed υ_x=c, then the phase velocity may be greater than the speed of light.

It follows from relation (6.34) that the phase velocity

$$\upsilon_f = \frac{\omega}{k} \tag{6.37}$$

The dependence of the phase velocity on the properties of the medium is called **dispersion**. It follows from (6.37) that the dispersion can be described by dispersion equations of the form

$$\left.\begin{array}{l} \omega = \omega(k) \\ k = k(\omega) \end{array}\right\} \tag{6.38}$$

$$\upsilon_f = \upsilon(\omega) \text{ and } \upsilon_f = \upsilon_f(k) \tag{6.39}$$

It is obvious that there is practically no dispersion of light in vacuum and homogeneous media. In this case, the harmonic waves

forming a complex wave (wave packet) propagate with the same phase velocity, and the packet behaves as a single, strictly localized entity. Under the influence of dispersion, packets of waves of different lengths are blurred as each monochromatic component of the packet moves with its own phase velocity.

The speed of the packet as a whole is called **group velocity**. It follows from this definition that the group velocity is the velocity of such a point in the wave at which the phase is the same for all wavelengths of the harmonic components, so the group velocity must be determined from the condition that

$$\frac{d\varphi}{dk} = 0 \tag{6.40}$$

$$\frac{d\varphi}{dk} = \frac{d\omega}{dk}t - x = 0 \tag{6.41}$$

from which

$$v_{\text{гр}} = \frac{x}{t} = \frac{d\omega}{dk} \tag{6.42}$$

Thus, in the absence of dispersion, the group velocity coincides with the phase velocity, and the wave is monochromatic.

6.5. Properties of waves

The properties of waves are specific and differ significantly from the properties of corpuscular bodies.

6.5.1. Transition of a wave from one medium to another

When a wave propagates in a homogeneous continuous medium without dissipation and with constant velocity, the magnitude of energy carried by it in a given direction does not change. However, at the boundary of different media, the wave undergoes a number of changes. These include reflection, dispersion, refraction, and absorption.

Reflection is a change in the direction of wave propagation at the boundary of two media without a transition from one medium to another. The direction of wave propagation is determined by the angle between the wave and the perpendicular to the boundary reconstructed from the point of wave incidence.

Dispersion is the reflection of a wave from different parts of a boundary at different angles.

Absorption occurs as a result of the interaction of the wave with the material medium, as a result of which part of the wave energy is absorbed by the particles of matter.

Refraction is the transition of a wave from one medium to another without absorption.

According to the Huygens principle, every point in the medium, which is reached by the disturbance, becomes the source of the secondary wave (Fig. 6.3). Thus, if all points of the medium are considered as point oscillators, the perturbation transmitted by the wave to the oscillator causes it to oscillate and it itself becomes the source of the secondary wave.

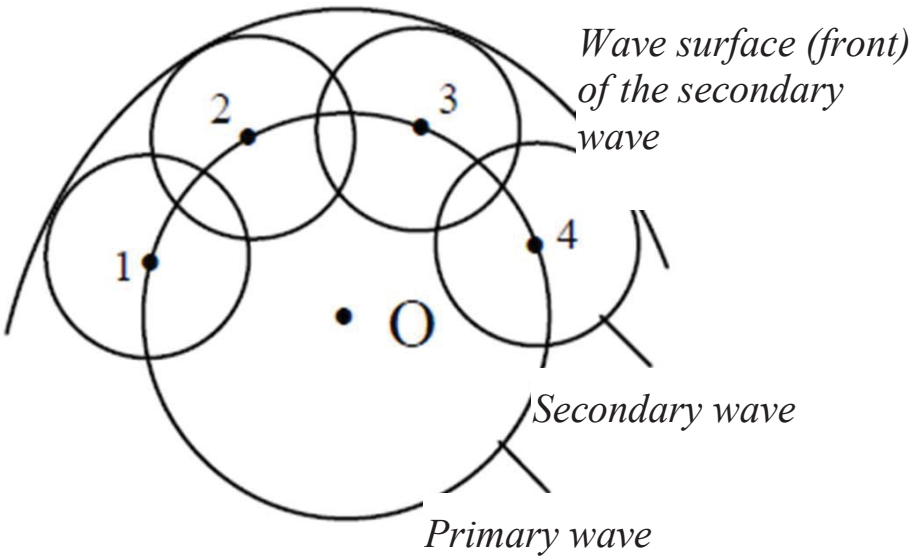


Figure 6.3.

Let a plane wave ray beam falls on the boundary surface *MN* of media 1 and 2 at an angle α (Fig. 6.4).

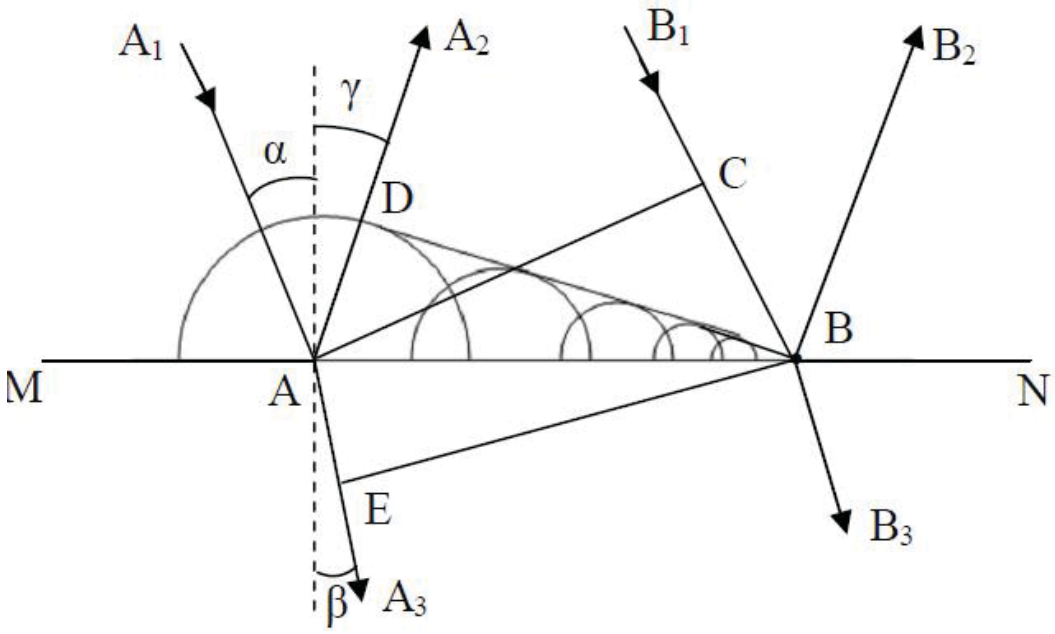


Figure 6.4.

The angle α is called the angle of incidence of the wave. All rays of this beam, as mentioned above, are parallel to each other. Let rays *A_1A and B_1B* be the outermost rays of the beam. According to the definition, AC is the front of the incident wave. According to the Huygens principle, all points of the AB segment of the interface become sources of secondary spherical waves, the envelope of which *DB* forms the front of the reflected wave.

This means that the reflected wave is formed by a bundle of parallel rays lying between its outermost rays *A_2A and B_2B*, perpendicular to its front DB. Obviously, the incident wave arrives at point B with some time lag relative to point A, equal to

$$\Delta t = \frac{CB}{\upsilon_1} \tag{6.43}$$

where υ_1 is the velocity of wave propagation in the medium 1.

At the moment when the wave arrives at point B, the secondary wave at point A is already a hemisphere with a radius of

$$AD = \upsilon\Delta t \tag{6.44}$$

Let us consider triangles *ADB* and *ACB*. It is not hard to see that

$$\Delta ADB = \Delta ACB \tag{6.45}$$

because they are (by definition) right triangles with a common hypotenuse AB and with equal legs AD = BC. From equality (6.45) it follows, therefore, that the angle of incidence α is equal to the angle of reflection γ. The result is called the law of reflection. Let us now assume that part of the incident wave passes from medium 1 to medium 2, having undergone refraction. We denote the angle of refraction by β*BE*. It is the front of the refracted wave, the outermost rays of which are AA_3 and BB_3. Let us consider *ΔACB* and *ΔAEB*. In these triangles

$$\mathrm{B}C = v_1 \Delta t \tag{6.46}$$

$$\mathrm{AE} = v_2 \Delta t \tag{6.47}$$

where v_2 is the speed of the refracted wave in the medium 2.

On the other side,

$$BC = AB sin\angle CAB$$

But $\angle CAB = \alpha$, as angles with respectively perpendicular sides, so

$$BC = AB \sin\alpha \tag{6.48}$$

For the same reason

$$AE = AB \sin\beta \tag{6.49}$$

We divide (6.48) by (6.49), then considering (6.46) and (6.47) we obtain that

$$\frac{\sin\alpha}{\sin\beta} = \frac{v_1}{v_2} = n \tag{6.50}$$

where *n* is a constant, independent of the angle of incidence α. The resulting equality is called Snell's **law of refraction**

Figure 6.4 shows that the incident, reflected, and refracted waves and the perpendicular reconstructed from the incident or reflection point

lie in the same plane. Therefore, the laws of reflection and refraction can be finally formulated as follows.

Law of reflection - the angle of incidence equals the angle of reflection, with the incident and reflected rays, as well as the perpendicular restored to the point of incidence, lying in the same plane.

The law of refraction - the ratio of the sines of the angles of incidence and refraction *(n)* is a constant, independent of the angle of incidence. This value is determined by the ratio of wave propagation velocities in the first and second media, respectively.

The value of the constant quantity *n* is called the **refractive index** of medium 2 with respect to medium 1. The refractive index of a medium relative to vacuum is called the **absolute** refractive index. From (6.50) in the case of a light wave we obtain for media 1 and 2, respectively

$$n_1 = \frac{c}{\upsilon_1} \; and \; n_2 = \frac{c}{\upsilon_2}$$

from which

$$\frac{n_1}{n_2} = \frac{\upsilon_2}{\upsilon_1} = n$$

The absolute refractive indices of the medium n_1 and n_2 are determined by the speed of wave propagation in the medium. The absolute refractive index of light characterizes the so-called **optical density of the medium**.

The larger it is, the lower the speed of wave propagation and vice versa. When a light wave passes from an optically less dense medium to an optically more dense medium (for example, from air to glass) $\upsilon_1 > \upsilon_2$ and, therefore, $n > 1$, and $\alpha > \beta$. This means that the refracted beam of light will approach the perpendicular. Otherwise, when, on the contrary, the beam passes from optically more dense medium to optically less dense medium $\beta > \alpha$, and the beam deviates more and more from the perpendicular and approaches the horizontal. The law of refraction in this case takes the form

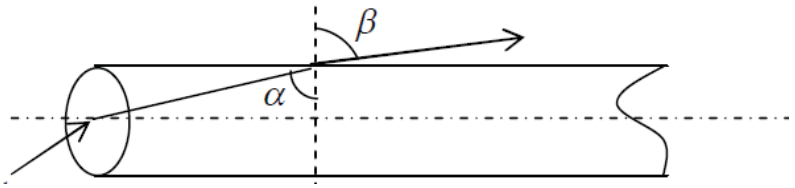
$$\frac{\sin\beta}{\sin\alpha} = n$$

At the critical value β = 90° and, consequently, α = α_{cr}, the refracted beam will go in the direction of the boundary and will not enter the second medium. Finally, if β > 90° and α > α_{cr}, the beam will return to the same medium, that is, it will actually be reflected from the interface. This phenomenon is called **total internal reflection**. In the critical regime, when β = 90° , i.e. with total internal reflection, the refractive law (6.50) takes the form

$$\left.\begin{aligned} \frac{\sin\beta}{\sin\alpha_0} &= n \\ \sin\alpha_0 &= \frac{1}{n} \end{aligned}\right\} \qquad (6.51)$$

The phenomenon of total internal reflection of light is used in fiber optics, which is a branch of physics that deals with the laws of light propagation in transparent media

Let us assume that a beam of light propagates inside a transparent light guide (see Fig. 6.5). Under the action of dispersion, it refracts at the boundary of the two media and changes the direction of its motion.



beam of light

Figure 6.5.

Bending the light guide also changes the direction of the beam. Let the ray entering the light guide fall on its inner surface at an angle α. Refracted, the ray leaves the limits of the guide and disperses. If the light guide surface is covered by a transparent cover with refractive index n_2, lower than that of glass n_1, or a light guide is made with variable refractive index, which gradually decreases from the axis to its surface, then, taking into account (6.51), a total internal reflection occurs and the ray, reflected many times from the cover or curved towards the axis, does not leave the light guide (Fig.6.6)

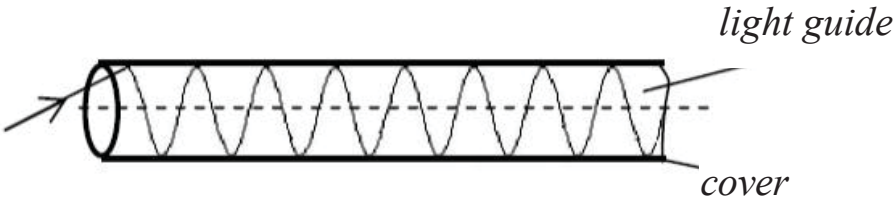


Figure 6.6.

In a curved light guide (see Fig. 6.7) the beam also does not go beyond the light guide and is not scattered, which allows the use of light guides to transmit light signals over long distances or to penetrate the vision in places difficult for visual inspection. Due to these properties, light guides have found wide application in communication engineering, medicine and many other branches of technology and in everyday life.

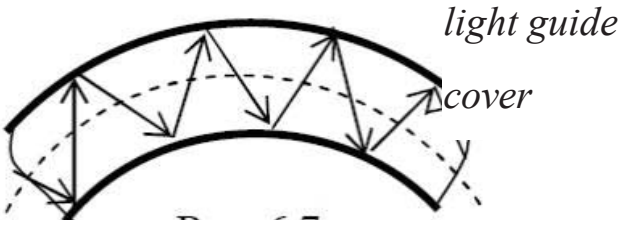


Figure 6.7.

6.5.2. Wave interference

When wave processes interact, the amplitude of the resulting wave depends in a complex way on the phase shift, amplitudes, periods, and lengths of the interacting waves. Let us consider, for example (Fig. 6.8), the result of adding two waves q_1 *and* q_2 with the same frequency, amplitude and length, but with opposite phases shifted by π/2 or by (2n+1) π /2, where n = 0,1,2, ... Then

$$\left.\begin{aligned} q_1 &= A\cos\varphi \\ q_2 &= A\cos\left(\varphi+\frac{\pi}{2}\right) \end{aligned}\right\} \qquad (6.52)$$

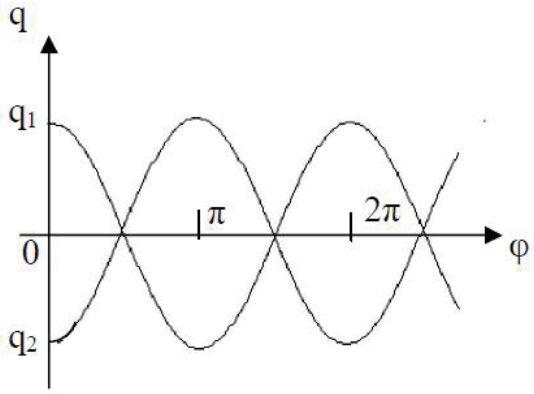


Figure 6.8.

This diagram shows that the resulting wave $q = q_1 + q_2 = 0$.

When adding two waves whose phase shift is zero or multiple of 2π, the resulting wave is equal (Fig. 6.9)

$$q = q_1 + q_2.$$

Thus, the two waves can mutually weaken and even cancel each other out as well as amplify

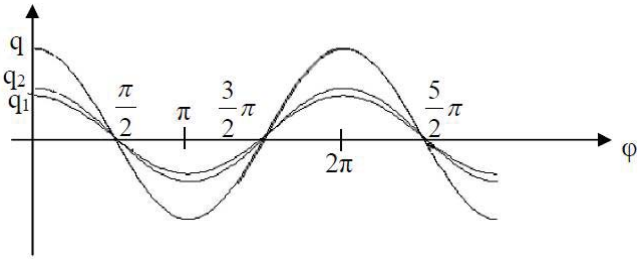


Figure 6.9.

Since light is not emitted simultaneously by different atoms, natural light is a mixture of light waves with different phases. When added together, these waves produce a resultant wave of light with some average amplitude. Directed at the screen, it illuminates it evenly with medium intensity. However, if we artificially provide a constant phase shift of two monochromatic waves of the same length, then, when added together, they strengthen or weaken each other. The screen will show maximum light at some points and minimum light at others. So, for example, if we direct two such rays, shifted in phase by π/2, to a point on the screen, instead of a light spot we get a dark spot. Waves with a constant (unchanging) phase shift in time are called **coherent**. The result

of adding coherent waves is called **interference**. The pattern of light distribution appearing on the screen is called an **interference pattern**.

The interference pattern is determined by the difference between the distances (travel difference) that the coherent waves travel from the sources to the screen. If these distances are denoted by d_1 *and* d_2, respectively, then the travel difference

$$\Delta d = d_1 - d_2 \tag{6.53}$$

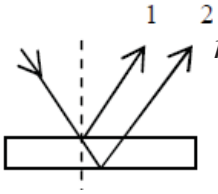
If the difference of path is equal to the wavelength A or a multiple of the wavelengths, then these waves, adding up, give maximums on the interference pattern. Thus, the condition for the appearance of **maxima**

$$\Delta d = k\lambda, \text{ where } k = 0, 1, 2, \ldots \tag{6.54}$$

If the difference is half the wavelength X/2 or an odd number of half-waves, the waves cancel each other out and give minima in the interference pattern. The condition for the occurrence of **minima** will be

$$\Delta d = (2k \pm 1)\frac{\lambda}{2}, where\ k = 0,1,2, \ldots \tag{6.55}$$

There are a number of ways to produce coherent waves. If, for example, a ray of light is sent to the surface of a transparent plane-parallel plate, part of the wave will be reflected from this surface, and another part, having refracted, will be reflected from the inner surface and, having refracted again, will leave the plate as a ray parallel and coherent to the original ray (Fig. 6.10).



rays 1 and 2 are coherent

Figure 6.10.

6.5.3. Wave diffraction

Waves, as the French physicist Augustin-Jean Fresnel first showed, propagate in a **straight line**. To justify this fact, Fresnel combined the Huygens principle with the idea of secondary wave interference. According to Fresnel's assumption, the wave surface at any moment of time is not simply an envelope of secondary waves, as follows from the Huygens principle, but is the result of their interference.

To calculate the amplitude of a wave at any point in space, it is sufficient to imagine a wave source surrounded by a closed surface and to calculate the interference of waves from secondary sources located on that surface.

Let a spherical wave from a point source S, reaches an arbitrary point B of space (Fig. 6.11).

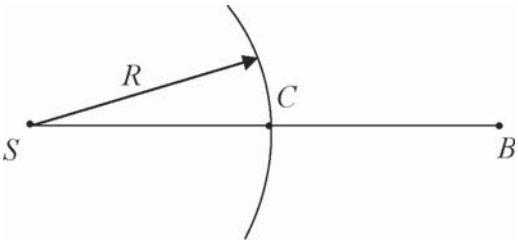


Figure 6.11.

Let's surround the source S with an arbitrary sphere surface of radius R. We connect points S and B with a straight line that intersects the spherical surface at point C.

Let us consider any secondary source on the surface located on one side of the straight line SB. Then there is always such a coherent source on the other side of it so that the waves from both sources cancel each other out. This means that only a wave of light propagating along the line CB from point C or from its small vicinity will arrive at point B.

The rectilinear propagation of light waves can be observed by looking at the shadow that appears behind any object in the path of the light emitted by the source (see Fig. 6.12). On the other hand, when a ray of light is emitted along the wall of a building, the observer who is around the corner of it does not see this ray (Fig. 6.13).

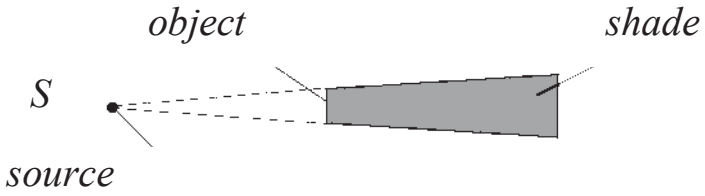


Figure 6.12.

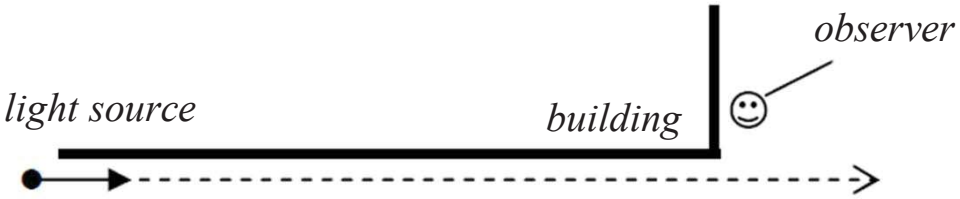


Figure 6.13.

This is not true for all types of waves. For example, a sound wave somehow penetrates around the corner of a building, and the sound emitted by the source is heard around the corner. This example suggests that waves can, under certain conditions, bend around obstacles and deviate from a straight-line propagation. The deviation of waves from a straight-line propagation is called **diffraction**. Diffraction of light was first observed by Thomas Young. An illustration of Young's experiment is shown in Figure 6.14.

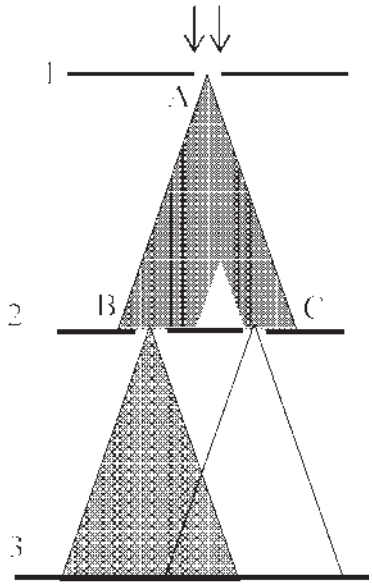


Figure 6.14.

A narrow beam of light passing through a small hole A in screen 1 illuminates screen 2, in which there are two small holes B and C. Passing through these holes, the light divides into two coherent beams, which overlap each other in a certain area.

As a result, the spherical wave coming from hole A excites secondary coherent waves in holes B and C. Due to diffraction, light comes out of B and C in two cones, which partially overlap and interfere. Therefore, an interference pattern in the form of alternating light and dark stripes appears on screen 3. When one of the holes is closed, the interference pattern disappears, which fully confirms Fresnel's theory.

The original optical device that allows to observe as well as to use the diffraction of light is the diffraction grating, which is a set of narrow slits cut out on an opaque base. The number of slits in the diffraction grating is very large. Usually they are located with a frequency of 3 - 4 thousand per 1 mm (Fig. 6.15).

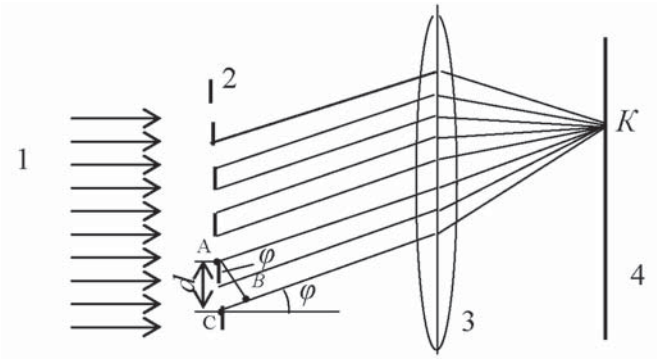


Figure 6.15.

Let us direct a parallel beam of light 1 of wavelength λ onto the diffraction grating. Secondary waves arising on the slits of grating 2 propagate in all directions. Let us choose one of them, given by the angle φ to the horizon. The difference of travel between two coherent waves coming from neighboring slits (see Fig. 6.15),

$$BC = d \sin \varphi \qquad (6.55)$$

Lens 3 is used to focus these waves at point K on screen 4. According to the condition of maxima

$$d \sin \varphi = k\lambda, \text{ where k = 0,1,2...,} \qquad (6.56)$$

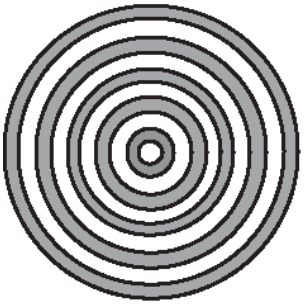


Figure 6.16.

A bright spot appears at point K. Obviously, there are a number of angles φ_i for which dark spots alternate with light spots at the corresponding points on the screen. Alternating light and dark rings appear on the screen from all the slits (Fig. 6.16). As follows from (6.56), the greater λ, the greater the distance between the maxima and the greater the deviation of the wave from straightness, its ability to skirt obstacles. In other words, the size of obstacles that are enveloped by waves as a result of diffraction must be of an order smaller than or equal to the wavelength. This explains why light waves of very short length cannot circle the building in the example above, while long wavelength sound waves easily circle it.

A diffraction grating can be used to make very precise measurements of the wavelength of light. If the grating period *d* is known, then, by measuring the angle φ_k indicating the direction to the *k*-th order maximum, we can determine λ from (6.56)

$$\lambda = \frac{d \sin \varphi_k}{k} \qquad (6.57)$$

6.5.4. Dispersion

Above (section 6.4) we defined wave dispersion as the dependence of wave speed on frequency or wave number (wavelength). On the other hand, considering the refraction of waves, we concluded that the speed of wave propagation in a medium is inversely proportional to the refractive index of that medium. It follows that the refractive index is

a function of wavelength, i.e. different monochromatic waves refract differently.

According to quantum theory, the speed of propagation of monochromatic light is proportional to wavelength. In other words, the speed of propagation of red light is greater than the speed of violet light. It follows that the red light beam is less refracted than the violet light beam. This fact was experimentally established even by Newton. He also suggested using the phenomenon of dispersion to decompose white light into its constituent monochromatic waves. The device proposed by Newton for this purpose turned out to be simple enough and contained a triangular glass prism, a screen, and a source of white light emitting beams of parallel rays (Fig. 6.17). Let us consider the course of one of these rays falling on the face of a prism in the direction parallel to its base AB. As the light wave passes from air to glass, it is refracted so that the beam approaches the perpendicular MN reconstructed to the face from the point of incidence C.

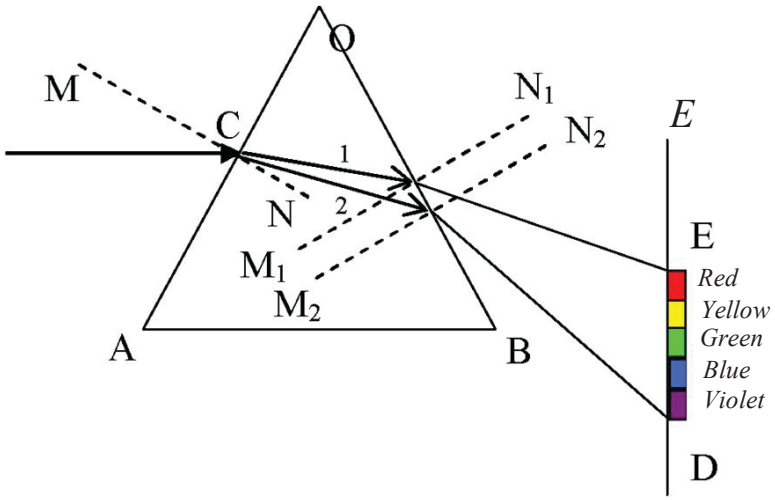


Figure 6.17.

Since white light contains the entire spectrum of colors of visible light (which was also experimentally proved by Newton), each of them is refracted differently. The strongest refraction is the violet ray 1, and the weakest is the red ray 2. Secondarily refracted on the second inner face, rays 1 and 2 diverge even more, and the whole spectrum of colors from red to violet appears on the *ED* screen.

The screen naturally has all the intermediate colors as well, so that the red color transitions smoothly into purple.

The phenomenon of decomposition of white light into its constituent monochromatic waves, each of which is perceived by its color, Newton called the **dispersion of light**, and the entire gamut of colors - a **continuum spectrum** of white light.

The dispersion of light explains the variety of colors in nature. Each object seems to be colored in the corresponding color because different substances have the property of reflecting some colors and absorbing others. For example, grass and tree leaves appear green to us because they reflect light with a wavelength that corresponds to the green color, while they absorb the rest. Sunlight scattering from particles in the upper atmosphere decomposes into its constituent colors (disperses) on these particles. The latter absorb light of all colors except blue, which they reflect. This is why the sky appears blue to us.

6.5.5. The Doppler effect

It is known from everyday experience that if the observer and the source of the sound wave move relative to each other or relative to the medium in which the wave process propagates, then the frequency of the wave (sound pitch) perceived by the observer differs from the frequency of the wave emitted by the source.

The dependence of the perceived frequency of a wave on the speeds of the source and the observer is called the **Doppler effect**.

In the case of mechanical waves, which carry medium perturbations, the Doppler effect is determined by the velocities of the observer and the source relative to the medium in which the wave propagates. In the case of non-mechanical waves, which do not carry a disturbance of the medium, but only a disturbance of some physical quantity, the Doppler effect is determined only by the relative speed of the source and the observer.

* For simplicity, let us assume that the observer M is connected to the stationary system K, and the wave source S is connected to the

moving system K'. Let us also assume that the source and the observer can move away from each other or get closer to each other.

In this connection, let us consider a conditionally stationary reference frame K and a frame K' whose axes are parallel to the axes of the frame K. Let the system K' move with respect to the system K progressively from left to right along the axis O with a constant velocity b (see Fig. 6.18a)

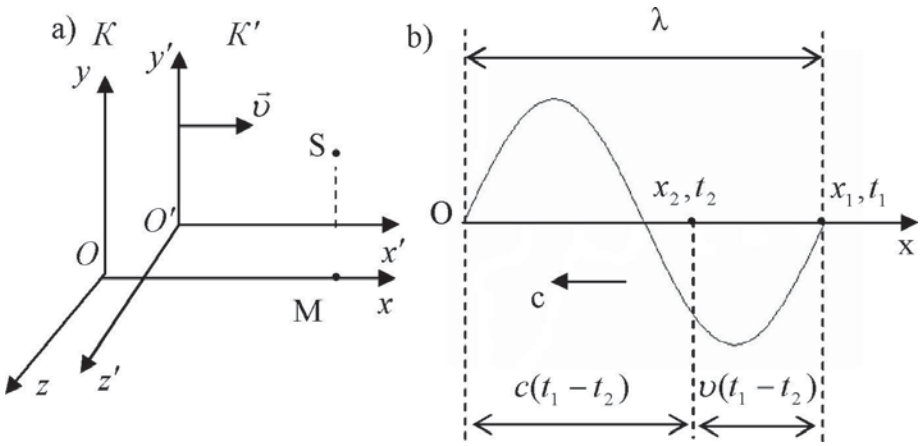


Figure 6.18.

Let, furthermore, the source of the wave S is moving away from the observer M. Let us assume that at the initial moment t_1 the source is at point x_1 of the system K (see Fig. 6.18b). In time (t_1 -t_2), the wave passes to the point (x_2,t_2), and propagates with speed c, passing in the system K a segment of path

$$c(t_1 - t_2).$$

At the same time, the system K' takes a path relative to the system K

$$\upsilon(t_1 - t_2).$$

Then (see Fig. 6.18b)

$$\lambda = c(t_1 - t_2) + \upsilon(t_1 - t_2) = (t_1 - t_2)(c - \upsilon) \qquad (6.58)$$

In the general case according to the Lorentz transformation in accordance with the theory of relativity (see Section 3.3)

$$t_1 - t_2 = \frac{(t_1' - t_2') + \frac{v(x_1' - x_2')}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{6.59}$$

The dashed values correspond to the coordinate and time in the K' system. Since the source is connected to the system K', then

$$x'_2 = x'_1, \text{ and } t'_1 - t'_2 = T$$

Substitution in (6.59) gives

$$t_1 - t_2 = \frac{T'}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{6.60}$$

Further, it is obvious that

$$\left.\begin{aligned} \lambda = cT = \frac{c}{v} \\ T' = \frac{1}{v'} \end{aligned}\right\} \tag{6.61}$$

where

v is the emitted frequency of the wave;

v' is the observed frequency.

Substituting (6.61) into (6.60) gives

$$v = v' \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}} \tag{6.62}$$

It follows from (6.62) that at the mutual distance of the source and the observer

$$v < v', \text{ and } \lambda > \lambda'$$

If the source and the observer approach each other, then the velocity b reverses its direction. To get the connection between v and v' in this case it is enough to change υ to (-υ) in the relation (6.62), and then

$$v = v' \sqrt{\frac{1 + \frac{\upsilon}{c}}{1 - \frac{\upsilon}{c}}} \tag{6.63}$$

v > v', and λ < λ'

Thus, the Doppler effect is reduced to a shift of the perceived frequency of the wave toward the long-wave part of the spectrum when the source and the observer are mutually distant. On the contrary, with mutual convergence - towards the shortwave part. It is easy to show that if the angle between the observer's line and the direction of the source relative to the observer is $\theta \neq 0$, then

$$v = v' \frac{\sqrt{1 - \frac{\upsilon^2}{c^2}}}{1 + \frac{\upsilon}{c} \cos\theta} \tag{6.64}$$

When $\theta = 0$, i.e., when the source and the observer are mutually distant, the relation (6.64) coincides with (6.62). At $\theta = \pi$, i.e. at their mutual convergence, the relation (6.64) coincides with (6.63).

The relations (6.62), (6.63) and (6.64) take into account the effect of relativity of space-time, which follows from the theory of relativity. Therefore, these formulas are suitable for calculating the impact of the Doppler effect in optics. As follows from these formulas, in optics, the Doppler effect becomes perceptible when the source and observer move relative to each other at a speed b close to the speed of light. When the speed b is far from the speed of light, the Doppler effect can only be detected with very sensitive and accurate instruments.

This is how the "red shift" (shift towards long waves) in the emission spectra of distant galaxies was discovered in the twenties of the 19th century, which was interpreted by the American astronomer Edwin

Powell Hubble as a consequence of these galaxies "escaping" from the observer. This was the first experimental confirmation of the expansion of the Universe.

In the case of mechanical (acoustic) waves at low velocities of their propagation, instead of the Lorentz transformation, the Galileo transformation applies. This transformation can be achieved by dividing the obtained relations by the relativistic coefficient

$$\sqrt{1-\frac{v^2}{c^2}}$$

In this case, the relation (6.64) takes the form

$$v = v'\frac{1}{1+\frac{v}{c}\cos\theta} \tag{6.65}$$

If we take into account that in the general case the source and the observer move with different velocities relative to the medium at different angles, then equation (6.65) takes a more complex form, which is given without proof

$$v = v'\frac{1-\frac{v_2}{c}\cos\theta_2}{1-\frac{v_1}{c}\cos\theta_1} \tag{6.66}$$

where υ_1 and υ_2 are the source and observer velocities, respectively relative to the medium;

θ_1 and θ_2 are, respectively, the angles between the vectors $\vec{v}_i$ и $\vec{k}_i$ at the moment in question.

6.5.6. Radiation

The process of propagation and transmission of waves created by oscillation sources from one point in space to another is called **radiation**.

Sources that excite waves are called **emitters**. Sound emitters, in particular, are devices designed to excite sound waves in various media. Sound emitters are divided into electroacoustic transducers, converters of energy of elastic vibrations into sound, gas-jet and hydrodynamic emitters, as well as musical instruments, noise devices, sound-producing apparatus of humans and animals, etc.

The emitters of electromagnetic waves are charges moving with acceleration, or bound charges (electrons bound in atoms), and on larger scales, antennas and other emitters.

Let us consider in more detail the theory of electromagnetic wave radiation. The simplest source of electromagnetic wave radiation is a point electric charge.

As follows from Maxwell's electromagnetic field theory, electromagnetic waves are emitted only by an accelerated moving charge. The instantaneous power of radiation N of such a charge is given without derivation

$$N = \frac{\mu_0 \mu^2}{6\pi c} a^2 \qquad (6.67)$$

Where

μ_0 is the magnetic constant (see Appendix 5);

μ is the magnetic permeability;

e is the magnitude of the charge;

a is the acceleration of the charge.

Radiation arising from the braking of a charge in a substance is called **braking radiation**. Radiation from a charge moving with acceleration in a magnetic field is called **synchrotron radiation**. If the charge performs harmonic oscillations, then in the one-dimensional version

$$x = A \sin \omega t \, ; a = \ddot{x} = -A\omega^2 \sin \omega t = -\omega^2 x \qquad (6.68)$$

Substituting (6.68) into (6.67) gives

$$N = \frac{\mu_0 \mu e^2}{6\pi c} \omega^4 x^2 \qquad (6.69)$$

Maxwell's theory shows that the simplest oscillatory system is an electric dipole with a shoulder l and a variable dipole moment d, which varies according to the periodic law

$$d = d_{max} \sin \omega t$$

In this case

$$N_{cp} = \frac{\mu_0}{6\pi c} \omega^4 d_{max}^2 \qquad (6.70)$$

Real emitters usually contain a large number of charges or many particles comprising these charges. However, those charges that are located away from the emitter have little or no effect on its emission. This makes it possible to replace the true charge distribution with an approximate one, and to replace the entire charge system with a single dipole, whose radiation is described by the relation (6.70). The theory of radiation proves that replacing a real emitter of electromagnetic waves with a radiating dipole is valid only if it is assumed that the arm of the dipole *l*

$$l << \lambda.$$

In engineering, the device for radiating (or receiving) electromagnetic waves is the **antenna** of a radio apparatus. The dipole antenna (see Fig. 6.19) was first constructed and used by Hertz. He achieved with its help the emission of electromagnetic waves in the range of lengths λ = 0.6-10 m, which corresponds to frequencies *ν = (0.5 - 0.03) ·10*^{3} MHz. The experimentally measured frequency of the wave emitted by the Hertz dipole antenna coincided with the obtained frequency from relation (6.70) with a fairly high accuracy.

Figure 6.19.

Thus, in the works of Hertz, Maxwell's theory was brilliantly confirmed.

Meanwhile, by the beginning of the 20th century, it became clear that classical electrodynamics, based on Maxwell's electromagnetic theory, was unable to explain the experimentally established patterns in the radiation spectra of atoms.

It turned out, in particular, that the radiation spectrum of atoms is not continuous, as it followed from the classical theory, but discrete, or, as they say, **line spectrum**. It was also found that the radiation of atoms is not described by the relation (6.67), but, at least for hydrogen-like atoms, by the experimentally derived Rydberg formula

$$\frac{1}{\lambda_{mn}} = RZ^2\left(\frac{1}{n^2} - \frac{1}{m_2}\right) \qquad (6.71)$$

where

λ_{mn} is the wavelength observed in the spectrum;

R is the Rydberg constant;

Z is the positive charge of the nucleus;

$m, n = 1, 2, 3, ...$are natural numbers.

As follows from (6.71), each given atom in the radiation spectrum corresponds to a well-defined line (or a set of lines) which always occupies the same place in the spectrum and has a coloring corresponding to a given wavelength λ_{mn}.

In the same way, the classical theory, based on Rutherford's so-called planetary model of the atom, could not explain the stability of atoms observed in practice. According to Rutherford's model, negatively

charged electrons rotate around the positively charged nucleus of the atom, the number of which is equal to the charge of the nucleus. According to Maxwell's theory (see 6.67), electrons, rotating around the nucleus of the atom, must continuously emit electromagnetic energy. This means that their emission spectrum must be continuous, and they themselves, continuously decreasing energy, must fall on the nucleus.

The resulting contradictions could only be resolved within the framework of quantum theory, the basics of which will be described in Chapter 7. Here we will only point out that according to quantum theory the energy of microparticles does not change continuously, but takes only certain values that form a discrete set.

Each of these values on the scale of energies occupies a different level. The transition of an atomic system from one energy level to another occurs in a stepwise fashion, bypassing the energy levels that are forbidden for it. In other words, an atom absorbs electromagnetic energy or emits it in discrete portions, quanta. Each such quantum of energy is equal to the value of

$$\mathrm{E} = \hbar\omega \qquad (6.72)$$

where

$\hbar$ is the Planck's constant;

ω is the cyclic radiation frequency of the atom.

$$\hbar = 1{,}05 \cdot 10^{-34}\mathrm{J} \cdot s$$

This quantum of electromagnetic energy is called a **photon**, which has both corpuscular and wave properties. This means that photons are characterized not only by frequency and wavelength, but also by mass and momentum. From this point of view, light is considered both as electromagnetic waves occupying a certain part of the radiation spectrum and as a set of photons emitted by atoms during their transition from a higher to a lower energy level.

Thus, according to quantum theory, all kinds of emission of electromagnetic waves are carried out by atoms. An atom is a system consisting of a positively charged nucleus and its associated negatively

charged electrons. Electrons are in one of the possible discrete states at their respective energy levels. The level with the lowest energy is called the **ground** level and is stable. In the absence of external influences, the electron can remain at this level for any length of time. All other permitted levels are unstable or **excitation** levels.

If an atom absorbs a quantum of energy (a photon), then under its influence it goes to one of the levels of excitation. Being here in an unstable state, the atom spontaneously, after some time, usually on the order of 10^{-8}s, returns to the ground state and emits the corresponding photon with energy

$$\mathcal{E}_i - \mathcal{E}_j = \omega\hbar$$

where i, j are the initial and final transition level.

The energy of the emitted photon thus depends on its frequency and is greater the greater the frequency. The energy of photons of the visible light spectrum, which are emitted by external electrons weakly bound to the nucleus, is relatively low. It increases as the bonding of the electron with the nucleus increases, i.e., the depth of the electron shell increases and the frequency increases (wavelength decreases). If upper shell electrons emit low-energy photons of visible light, then inner shell electrons emit high-energy photons of X-rays and emission of hard gamma rays (see also clause 6.6.3)

Quantum theory proves that the radiation intensity of an atom N_{ij} (the intensity of lines in the spectrum) is determined by a formula close to the classical relation (6.70), namely

$$N_{ij} = \frac{\mu_0 \mu \omega_{ij}^4}{6\pi c} d_{ij}^2 \qquad (6.73)$$

The value of d_{ij} is the quantized analogue of the dipole moment.

In addition to the spontaneous radiation of the atom, there is also stimulated radiation, which occurs when the atom absorbs an external photon that coincides in frequency with one of the frequencies allowed for it. In this case it is not in the ground, but in the excited state. The atom emits a photon of the same frequency as the absorbed photon. The

probability of stimulated emission, unlike spontaneous emission, is proportional not to frequency, but to the intensity of absorbed photons, i.e. their number.

6.5.7. Poynting Vector. Polarization

Non-mechanical waves, as mentioned above, include electromagnetic and hypothetical gravitational waves. The properties of the latter, discovered recently (in 2015), are being studied. The concept of electromagnetic waves first appeared in Maxwell's electromagnetic theory in the middle of the 19th century. At the end of the 19th century, electromagnetic waves were discovered in very subtle physical experiments by Hertz. When considering Maxwell's equations, it was found (Section 4.2.3.3) that their solutions are described by the vector and scalar potentials of the electromagnetic field $\vec{A}$ and φ, which are given by equations (4.173). Each of the four roots of these equations (A_x, A_y, A_z, φ) satisfies the d'Alembert's formula (4.174). If we consider the electromagnetic field in a nonconductive (immaterial) medium, for example in a vacuum, where there are no currents and free charges ($\vec{j}$ = 0, ρ = 0), then the d'Alembert's formula takes the form (see equation 4.177)

$$\nabla^2 S_i - \frac{1\delta^2 S_i}{c^2 \delta t^2} = 0 \qquad (6.74)$$

where S_i is one of the four solutions of the d'Alembert's formulas, called the electromagnetic wave;

c is the electrodynamic constant

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \qquad (6.75)$$

Substituting the values of the electrical and magnetic constant (see Appendix 5) into (6.85) gives

$$c = 3 \cdot 10^8 \frac{m}{s} \tag{6.76}$$

From (6.74) and (6.76) we see that the value c is the speed of propagation of the wave process of the electromagnetic field in the vacuum. When a wave propagates in matter, the value of its speed of propagation is

$$\upsilon = \frac{1}{\sqrt{\varepsilon_0 \mu_0 \varepsilon \mu}} \tag{6.77}$$

Since $\varepsilon > 1$ и $\mu > 1$, the speed of propagation of electromagnetic waves in matter

$$\upsilon < c \tag{6.78}$$

In other words, the value of c is the maximum possible speed of propagation of electromagnetic field waves.

Since the potentials A_i and φ characterize, respectively, the magnetic field $\vec{H}(t)$ and the electric field $\vec{E}(t)$, it follows from the d'Alembert's formula (6.74) that these fields must be aligned for an electromagnetic wave to occur. Considering an oscillating circuit, we made sure that the oscillating fields $\vec{E}(t)$ and $\vec{H}(t)$, arising in the circuit, are disconnected. The field $\vec{E}(t)$ is concentrated in the capacitor, and the field $\vec{H}(t)$ is concentrated in the loop inductance coil. To combine these fields, as Hertz first showed, the oscillating circuit must be opened (see Fig. 6.20 a,b,c)

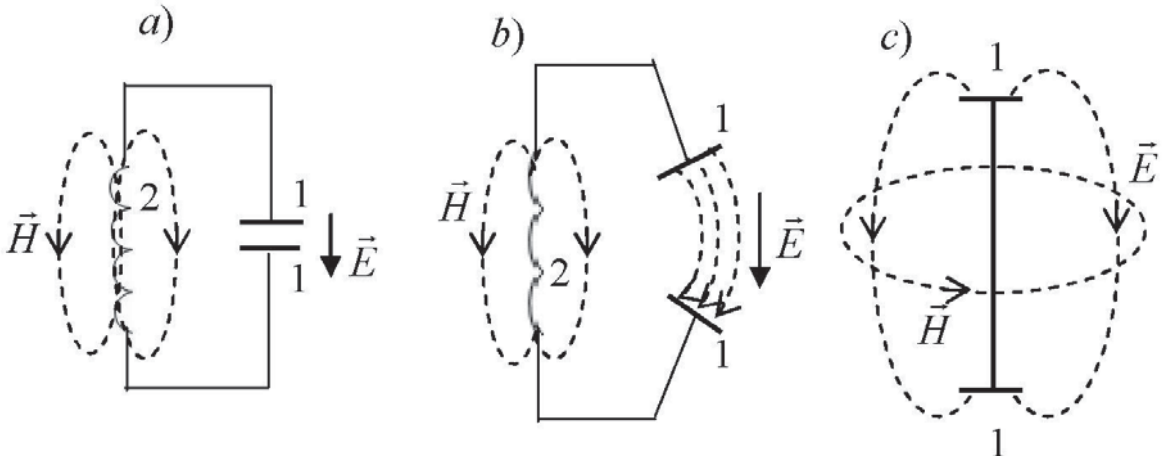


Figure 6.20.

The force lines of the fields $\vec{E}$ and $\vec{H}$, in the open loop are arranged in such a way that these fields coincide in space-time (see Fig. 6.20c).

This results in an electromagnetic wave that bounces off the circuit under certain conditions.

The solution of the d'Alembert's formula, as shown above (see Equation 4.180), are the so-called delayed potentials, which, as follows from the above equation, indicate that the electromagnetic wave propagates in time with a finite speed from one point of space to another.

It also follows from the d'Alembert's formula (6.74) that the electromagnetic field potentials satisfy the wave equation. This, in turn, means that the electric and magnetic field strengths uniquely associated with the potentials must also satisfy the corresponding wave equations. Thus, in particular, for the strength E by analogy with (6.74) in one-dimensional approximation we can write

$$\frac{\partial^2 E}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = 0 \qquad (6.79)$$

The solution of equation (6.79) according to (6.12) is

$$E = \psi(x - ct) \qquad (6.80)$$

On the other hand, when solving Problem 2, Section 4.2.3.5, it is obtained that the energy Ɛ of the electromagnetic field

$$\mathcal{E} = \frac{1}{2}\int_v \varepsilon_0 \varepsilon E^2 dV + \frac{1}{2}\int_v \mu_0 \mu H^2 dV \tag{6.81}$$

The differentiation of (6.81) gives

$$\frac{\partial \mathcal{E}}{\partial t} = \int_v \left[\varepsilon_0 \varepsilon \left(\vec{E}\frac{\partial \vec{E}}{\partial t}\right) + \mu_0 \mu \left(\vec{H}\frac{\partial \vec{H}}{\partial t}\right)\right] dV \tag{6.82}$$

In a non-conductive medium, where $\vec{j} = 0$, Maxwell's 1st and 2nd equations, (see 4.160), are written in the form

$$\left.\begin{aligned} rot\vec{E} &= -\mu_0 \mu \frac{\partial \vec{H}}{\partial t} \\ rot\vec{H} &= \varepsilon_0 \varepsilon \frac{\partial \vec{E}}{\partial t} \end{aligned}\right\} \tag{6.83}$$

Substitution of $\frac{\partial \vec{E}}{\partial t}$ и $\frac{\partial \vec{H}}{\partial t}$ from (6.83) into equation (6.82) gives

$$\frac{\partial \mathcal{E}}{\partial t} = \int_v \varepsilon_0 \varepsilon \left(\vec{E}, \frac{rot\vec{H}}{\varepsilon_0 \varepsilon}\right) dV - \int_v \mu_0 \mu \left(\vec{H}, \frac{1}{\mu_0 \mu} rot\vec{E}\right) dV$$

From the equations of vector analysis (see Appendix 3) it follows that

$$\left(\vec{b}, rot\vec{a}\right) - \left(\vec{a}, rot\vec{b}\right) = div\left[\vec{a}, \vec{b}\right] \tag{6.84}$$

and

$$\int_v div\left[\vec{a}, \vec{b}\right] dV = \oint_s \left(\left[\vec{a}, \vec{b}\right], d\vec{S}\right) \tag{6.85}$$

Applying (6.84) to (6.83), we obtain that

$$\frac{\partial \mathcal{E}}{\partial t} = -\int_{v} div\left[\vec{E}, \vec{H}\right] dV \tag{6.86}$$

or, according to (6.85)

$$\frac{\partial \mathcal{E}}{\partial t} = -\oint_{s} \left(\left[\vec{E}, \vec{H}\right], \vec{n}_0\right) dS \tag{6.87}$$

When propagating, an electromagnetic wave carries with it a flux of electromagnetic energy Ɛ. *

Let us consider an arbitrary closed surface $\vec{S}$, within which a wave propagates. The wave energy per elementary area dS in time dt is called the energy flux density Y. Let us introduce the energy flux density vector $\vec{Y}$. According to the definition

$$dY = \frac{\partial \mathcal{E}}{\partial t dS} \cdot \vec{n}_0 \tag{6.88}$$

where $\vec{n}_0$ is the unit vector of the area $d\vec{S}$.

From (6.87) we obtain that

$$\frac{\partial \mathcal{E}}{\partial t dS} \cdot \vec{n}_0 dS = \left[\vec{E}, \vec{H}\right] dS, \qquad \vec{Y} = \left[\vec{E}, \vec{H}\right] \tag{6.89}$$

Since the vector $\vec{Y}$ determines the direction of the energy flux, it coincides with the direction of propagation of the electromagnetic wave.

It follows from equation (6.89) that the vectors E and H of an electromagnetic wave form a right-handed orthogonal bundle with the direction of its propagation n. To explain, let us introduce a Cartesian coordinate system in which the Ox axis for a one-dimensional (plane) electromagnetic wave coincides with the direction of its propagation.

Then the Oy axis will coincide with the direction of vector E, and the Oz axis with the direction of vector H (see Fig. 6.22)

* This means that

$$E_x = E_Z = 0,$$

$$E_\gamma = E$$

and

$$\frac{\partial \mathcal{E}}{\partial x} \neq 0 \left(\frac{\partial \mathcal{E}}{\partial y} = \frac{\partial \mathcal{E}}{\partial z} = 0\right)$$

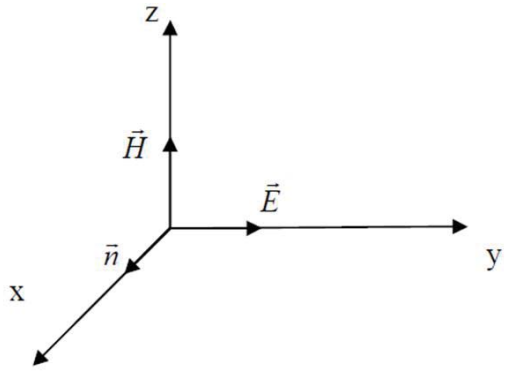


Figure 6.21.

Since according to Maxwell's second equation

$$rot\vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{6.90}$$

then

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t} \tag{6.91}$$

Let's substitute equation (6.80) in (6.91), then

$$\frac{\partial B}{\partial t} = -\frac{\partial \omega(x-ct)}{\partial (x-ct)} \cdot \frac{\partial (x-ct)}{\partial x} = -\frac{\partial \psi}{\partial x} \tag{6.92}$$

from which

$$B = -\int \frac{\partial \psi}{\partial x} dt = \frac{1}{c}\int \frac{\partial \psi}{\partial x} d(x - ct)$$
$$= \frac{1}{c}\int \frac{\partial}{\partial x}(x - ct) d(x - ct) \qquad (6.93)$$
$$= \frac{\psi(x - ct)}{c} + F(x)$$

F(x) is an arbitrary integration function which can be determined from the boundary conditions for induction B, namely, assuming that when E ≠ 0, B ≠ 0 , and when E = 0, B = 0.

It follows that F(x) = 0.

Thus, from (6.93) we obtain that

$$B = \frac{E}{c} \text{ or } (B = \frac{E}{\upsilon} \text{ in the matter})$$

Since

$$\upsilon = \frac{1}{\sqrt{\varepsilon_0 \varepsilon \mu_0 \mu}} \qquad (6.94)$$

then

$$B = E\sqrt{\varepsilon_0 \varepsilon \mu_0 \mu}$$

and

$$H = E\sqrt{\frac{\varepsilon_0 \varepsilon}{\mu_0 \mu}} \qquad (6.95)$$

It follows from relation (6.94) that electric and magnetic fields in a plane one-dimensional wave are always in-phase so that their characteristics change in time and space quite equally.

Energy flux density vector

$$\vec{Y} = [\vec{E}, \vec{H}] \qquad (6.96)$$

is called **the Poynting vector.**

From equation (6.96) there also follows another very important conclusion, according to which electromagnetic waves are transverse. This was proven experimentally by the observation of a phenomenon called **wave polarization**. In the general case, in the process of wave propagation, vectors $\vec{E}$ и $\vec{H}$ can change their directions, but along with them should change their orientation and the plane in which they oscillate so that the right-hand orthogonal coupling of these vectors with the direction of wave propagation is not violated. The plane in which a vector $\vec{E}$ и $\vec{H}$ in a given monochromatic wave vibrates is called the **polarization plane**. Since any real wave consists of many elementary waves arising under different conditions and at different times, it contains in its composition a bundle of many elementary wave impulses, each of which has its own polarization planes. It turned out, however, that there are ways in which you can get a wave with a fixed plane of polarization. Such a wave is called a **polarized wave**.

6.6. Types of wave processes

It was already mentioned above that depending on the specifics of the propagating disturbances, waves are divided into mechanical, non-mechanical, and matter waves according to the conception of wave-particle duality (de Broglie waves). In addition, waves are divided into standing and running waves, wave trains, and solitons.

The following are examples of typical wave processes.

6.6.1. Standing waves

Let us consider the result of interference of coherent waves of the same frequency and intensity propagating in mutually opposite directions. To this end, for example, let's imagine a running wave on a perfectly reflecting surface perpendicular to that surface. In this case, propagating towards the incident wave, the reflected wave will be coherent to it. Since the reflecting surface is assumed to be ideal, the intensity of the reflected wave is equal to the intensity of the incident wave.

Let the plane incident wave in the one-dimensional version is defined by the wave function

$$\psi_1(x,t) = A\cos(\omega t - kx + \varphi_{01}) \tag{6.97}$$

where φ_{01} is the initial phase of the wave.

Then the reflected wave, subject to the above conditions, has the form

$$\psi_2(x,t) = A\cos(\omega t + kx + \varphi_{02}) \tag{6.98}$$

The addition of the wave functions of the incident and reflected waves ψ_1 and ψ_2 results in a new wave

$$\psi(x,t) = A[\cos(\omega t - kx + \varphi_{01}) + \cos(\omega t + kx + \varphi_{02})]$$

or

$$\psi(x,t) = 2A\cos\left(kx + \frac{\varphi_{02} - \varphi_{01}}{2}\right)\cos\left(\omega t + \frac{\varphi_{01} + \varphi_{02}}{2}\right) \quad (6.99)$$

The resulting wave, according to (6.99), is an oscillatory function of time with a periodically varying amplitude in space, which does not depend on time. Thus, the newly formed wave as if frozen in space, stops. That's why it's called **standing**.

We denote the amplitude of the standing wave A_{st}. It follows from (6.99) that

$$A_{st} = 2A\cos\left(kx + \frac{\varphi_{02} - \varphi_{01}}{2}\right) \quad (6.100)$$

The points of space (values of the x-coordinate) at which

$$A_{st} = 0, \quad (6.101)$$

are called standing wave **nodes**, and the points at which A_{st} has a maximum value, 2A, are called **antinodes**.

It follows from (6.100) that for the nodes

$$kx + \frac{\varphi_{02} - \varphi_{01}}{2} = (2m-1)\frac{\pi}{2} \quad (6.102)$$

and for antinodes

$$kx + \frac{\varphi_{02} - \varphi_{01}}{2} = 2_m\frac{\pi}{2} \quad (6.103)$$

In the latter formulas m = 0, ± 1, ±2, ±3, ±...

The standing wave length λ_{st} is the **distance** between two adjacent nodes or antinodes. According to the definition, the length of the initial traveling wave $\lambda_{st} = \lambda/2$. The distance between the neighboring node and the antinode is $\lambda/4$.

In a standing wave, all particles of the medium between two neighboring nodes (antinodes) oscillate in the same phase, but with

different amplitudes. When passing through a node, the phase of the oscillation changes by n as the sign of the function $\psi(k)$ changes.

Unlike traveling waves, standing waves do not transfer energy. In them, only the spatial transfer of energy of one kind to another is carried out. In this sense, the vibrations of elastic bodies can be seen as standing waves in these bodies.

An example of standing waves is a plane sound wave inside an air- or liquid-filled tube when the ends are closed or open. On the solid wall, for obvious reasons, a velocity node and a differential pressure antinode of the medium are formed. On the open end, on the contrary, there is a velocity antinode and a pressure drop node. Therefore, the patterns of pressure and velocity of standing waves are shifted relative to each other by a quarter of the wavelength (see Fig. 6.22)

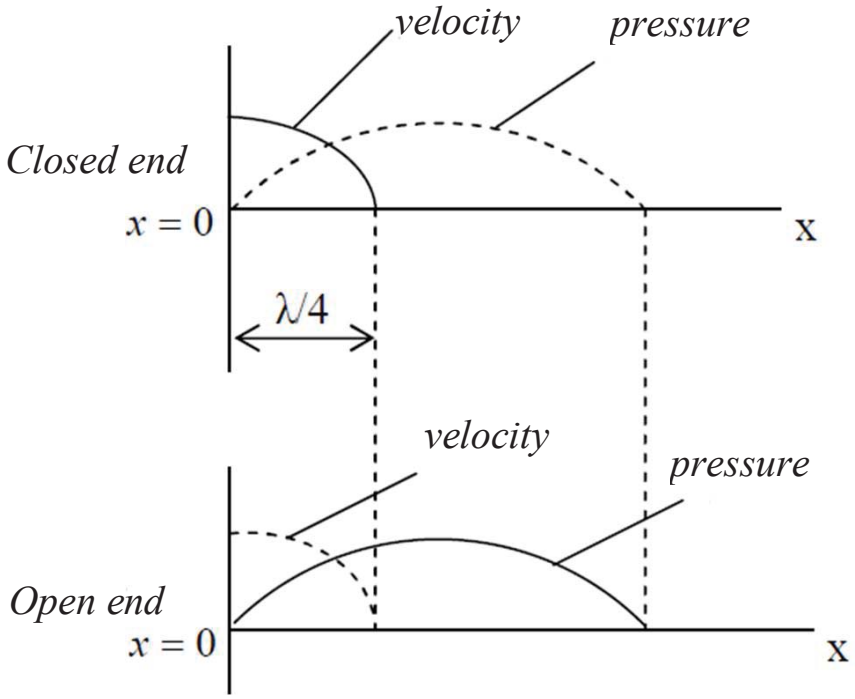


Figure 6.22.

Another example of standing waves can be the distribution of electric and magnetic fields in a transmission line or waveguide with a perfect closed or open end.

If a liquid or gas fills a limited portion of the space in a vessel, a complex system of standing waves arises, which depends on the shape

and size of the vessel. The frequencies of natural vibrations of these waves are determined by their length.

Thus, for example, in a column of gas or liquid in cylindrical vessels, standing waves are formed with the following series of natural frequencies:

1) if both ends of the vessel are closed, then according to (6.103), assuming that the oscillations are in-phase,

$$kx = 2m\frac{\pi}{2}$$

At $x = l$, where l is the length of the pipe,

$$kl = m\pi;$$

$$\frac{2\pi}{\lambda} l = m\pi$$

$$\frac{2l}{c} v_m = m$$

$$v_m = \frac{mc}{2l} \qquad (6.104)$$

2) if one end of the pipe is closed and the other end is open, then according to (6.102) under the same conditions

$$kx = (2m-1)\frac{\pi}{2}$$

$$\frac{2\pi}{\lambda} l = (2m-1)\frac{\pi}{2}$$

$$\frac{2+1}{c} v_m = (2m-1)\frac{1}{2}$$

$$v_m = \frac{(2m-1)c\frac{1}{2}}{2l}$$

$$\nu_m = \frac{(2m-1)c}{4l} \tag{6.105}$$

Let us consider an oscillating system capable of oscillating with maximum amplitude under the influence of external periodic factors. Such a system is called a **resonator**. The oscillations inside the resonator, as in any elastic body, are distributed in the form of standing waves. It can be shown that the minimum natural frequency of resonators (m = 1) with a long thin tube, by which they communicate with the external environment, will be

$$\nu_m = \nu_{min} = \frac{c}{2\pi}\sqrt{\frac{S}{lV}} \tag{6.106}$$

where

V is the volume of the vessel (volume of the cavity of the resonator);

S is the cross-sectional area of the tube;

l is the length of the tube.

Let's further consider free standing waves using the example of a sounding stretched string.

If you fasten a string of length l at its ends, when you excite oscillations in it, a standing wave with two nodes (at the fastened ends) and one antinode (between the nodes) appears. If you fasten the string at one more point (for example, press it against a surface with your finger), a standing wave with three nodes and two antinodes will arise in the string. In this case, the parts of the string between two adjacent nodes can be considered as two strings. If the string is pressed in the middle, the length of each section is $l/2$, and the natural frequency of the standing wave, according to (6.104), is doubled. It is possible to excite 3, 4, 5, etc. natural vibrations in a string, the frequency of each of which increases by a factor of 3, 4, 5, etc., i.e. by a multiple of the minimum frequency. Thus,

a whole set of standing waves with multiples of natural frequencies can be excited in the string.

A set of standing waves can be excited in an extended object of any shape. This is exactly what happens in the resonators already mentioned above. The lowest frequency of oscillations excited in such objects (strings, boxes, pipes, resonators, etc.) is called the **fundamental frequency**. Oscillations with the fundamental frequency are called the **fundamental tone**, and oscillations with multiples of the fundamental tone frequencies are called overtones, first, second, etc.

A perfect standing wave can only be established in the absence of attenuation (absorption) and refraction on reflective surfaces.

Otherwise, along with standing waves, traveling waves are also formed, delivering energy to the places of absorption and radiation.

It also follows from the above that at the distance limiting the standing wave region, only such standing waves can be formed in which an integer number of half-waves are stacked at length l. If this condition is not met, the interference of counter waves produces a non-harmonic process with irregularly changing amplitudes and phases. An oscillating circuit is an example of a standing electromagnetic wave generator. When the electromagnetic energy of a circuit is concentrated in an inductor coil, it is represented as an antinode of magnetic energy, and when it is in a capacitor, it is represented as an antinode of electrical energy. Accordingly, in a standing wave there is a reciprocating movement of energy between the magnetic and electric antinodes. Meanwhile, in a travelling wave there is no mutual transformation of electric and magnetic energy, but both types of energy are simply transported in space.

6.6.2. Solitons

It was shown above (section 6.4) that the vast majority of oscillatory wave processes common in Nature or created by human creative activity are characterized by small amplitude, are described by harmonic sine or cosine functions and are therefore linear. These include, for example, harmonic waves of sound, light, presumably gravitational waves, etc., as well as acoustic waves, electromagnetic waves, including radio waves and light. Nonlinear waves, i.e. waves with increased amplitude, are much less common in Nature. Solitons are an example of such waves, as has been shown (see Section 6.4). Solitons are wave trains whose crests degenerate into one structurally stable crest in the limit.

A soliton, in fact, is not a single wave, but a spectrum of waves. However, unlike the usual wave train, the soliton includes a continuous spectrum of an infinite number of harmonic waves, the envelope of which is pulled down to a point and is perceived as a single wave. In the case of soliton propagation in a homogeneous non-dispersive medium, it is, due to its nonlinearity, very unstable and quickly collapses. The point is that nonlinear waves differ from linear waves in that the velocity of their different points is different, and as the point approaches the crest its velocity increases. This initially leads to an increase in the steepness of the wave leading edge and to its self-compression, but after a while the crest tip overturns and the soliton is destroyed. The soliton behaves differently in a dispersive medium. Under the action of dispersion, monochromatic waves of different lengths forming the wave train also move with different speeds, but in such a way that it leads to a decrease in the steepness of the wave fronts and the blurring of the wave train, which compensates for its self-compression at a certain dispersion.

A soliton is a wave. Like any wave, it does not carry matter, but energy. Its energy, like all waves, is determined by the square of the amplitude; theoretically, it can be decomposed into its constituent monochromatic waves. At the same time, it is very similar to a particle of matter. Like any particle, it is quite well localized, spatially stable, behaves like a particle as a unit, moves like a particle with a small speed, interacts with other solitons not as a wave, but as a particle, repulsing

from them. It, as a particle, cannot quench another soliton or amplify it, it cannot go around obstacles (it has no wave properties of interference and diffraction).

On the basis of the formal similarity, the soliton is often confused with the photon, although they are completely different objects. A photon, unlike a soliton, is not a packet of an infinite number of monochromatic waves, but a single monochromatic wave with a specific frequency and wavelength. A free photon is not localized like a soliton, but, on the contrary, is distributed throughout infinite space-time. Spectrally, it appears as an infinitely thin line, while the spectrum of the soliton is solid.

The photon exhibits all wave properties, while the soliton lacks most of them. Photon propagates in any medium, including in vacuum, while soliton cannot propagate in a homogeneous, non-dispersive medium, certainly not in vacuum, etc.

Solitons can be of different physical nature. It has been proved, for example, that dislocations (holes), which take part in the electrical conductivity of solids, are solitons, i.e., elastic waves. Nerve impulses in the organisms of higher animals are solitons in nature. Although, unlike light waves, solitons cannot propagate in a vacuum, there are also optical solitons, which have recently found applications in communication engineering. It has already been mentioned above that light signals can propagate by means of light guides over long distances. However, in the process of movement in the light pipe, they still gradually dissipate so that after every 80-100 km they fade out and have to be restored with the help of expensive retransmission equipment. It turned out that if the light signal is transmitted by solitons, the distance between the retransmission centers can be up to 5-6 thousand kilometers. For this, however, the light guide must be made with controlled dispersion. Currently, solitons have found applications in many other fields of engineering and medicine.

6.6.3.Electromagnetic waves. Light waves

The properties of wave processes are largely determined by their length. The totality of the harmonic components of waves in a given range of lengths is called the wave **spectrum**. Above we have already considered the electromagnetic waves arising from the emission of atoms belonging to different parts of the spectrum.

The entire spectrum of electromagnetic waves occupies a wide range that extends from 10^5 to 10^{-15} m. Long wavelengths from 10^5 to 10^{-4} m form the spectrum of radio waves, which are widely used in engineering. Beyond radio waves there are optical spectrum waves, X-rays, hard γ-rays, etc. Figure 6.23 shows the scale of electromagnetic waves on a logarithmic scale.

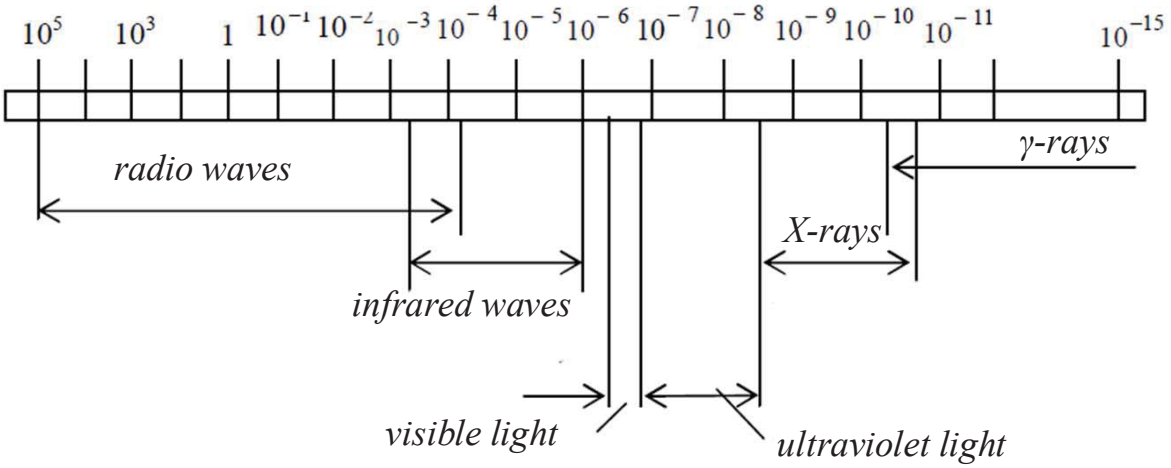


Figure 6.23.

Conventionally, all types of electromagnetic radiation are called light. Visible light, as seen in Figure 6.23, occupies a very narrow spectrum between 0.45 and 0.75 μm. To the left and right of the visible light spectrum are the broader invisible spectra of infrared and ultraviolet light.

It has already been mentioned above that according to electromagnetic theory, the speed of propagation of electromagnetic waves in a vacuum is equal to the speed of light. From this Maxwell concluded that light is inherently electromagnetic waves. The wave nature of light was proved earlier by Huygens and Fresnel. They

experimentally showed that light has all the known wave properties - dispersion, interference, diffraction, polarization, etc. However, by analogy with sound, light was considered as longitudinal waves propagating in a special elastic medium, the ether. Maxwell's theory confirmed the previously established transverse nature of light waves. Hence the conclusion that light must have the property of polarization. However, for a long time it was not possible to detect the polarization of light. The fact is that light is emitted by natural sources and is, therefore, a mixture of waves with different polarizations. It turned out, however, that there are some crystals that are able to transmit light waves, if they are polarized, only with a certain polarization direction. This direction is called the optical axis of the crystal. Such crystals include tourmaline in particular. This fact made it possible to detect the polarization of light experimentally by passing it through a plate of tourmaline. If light really has the property of polarization, then the plate of tourmaline must have transmitted it only in the direction of the optical axis. To detect polarization in this case, it was enough to pass the polarized light through a similar second plate of tourmaline. By rotating it relative to the first plate, we could see that in a certain orientation, when its optical axis was perpendicular to the optical axis of the first plate, it completely blocked the light. This means that it did not allow the light polarized by the first plate to pass in the direction of the optical axis of the tourmaline. On the contrary, when the optical axes of both plates are parallel, the light is passed through the second plate. This experiment confirmed the transverse nature of light waves and eliminated doubts about their electromagnetic nature.

Humans have pondered the nature of light throughout their existence. The rectilinear distribution of light, for example, was already known to the peoples of the Middle East at least 5,000 years BC and was used in ancient Egypt for construction. Pythagoras (6th century BC) was one of the founders of the corpuscular theory of light. Aristotle (4th century BC) made an ingenious guess close to modern wave theory, believing that light is the result of excitation of the medium between the visible object and the human eye. He was also not far from the truth when he viewed the rainbow as the result of the reflection of light by water

droplets. The rectilinearity of light and the laws of reflection were first formulated by Plato's school. Euclid was the founder of geometric optics and studied the processes of the emergence of reflection in a mirror. The refraction of light and the path of rays in lenses were studied in the Middle Ages. The first optical instruments appeared in the 13th century. The first telescope was built by Galileo in 1609. The laws of refraction were first formulated by Willebrord Snellius in 1620 and by Descartes in 1637. Diffraction of light was discovered by the Italian scientist Francesco Maria Grimaldi and results of his discovery was published in 1665, and interference of light was discovered in 1669 by the Danish scientist Rasmus Bartholin. One of the pillars of the Renaissance, the Italian scientist and artist Leonardo da Vinci, and the English scientist Robert Hooke, who adhered to the wave theory, were also concerned with the nature of light. This theory was further developed in the works of Huygens, who, like his predecessors, proceeded from the analogy of light and sound. Newton assumed the interpretation of light as a wave process, but gave preference to the corpuscular theory, considering light as a stream of particles acting on ether, the concept of which as an elastic light-carrying medium was introduced by Descartes. Kepler and Descartes considered the speed of light to be infinite. The notion of the finiteness of the speed of light was introduced by Hooke and Newton. The speed of light was first measured experimentally in 1676 by the Danish astronomer Ole Rømer. Despite Huygens' convincing work on the wave properties of light, the corpuscular theory, based on Newton's authority, lasted until the beginning of the 19th century. The final blow was dealt to it by the work of the English scientist Young and the French scientist Fresnel.

In 1801 the laws of interference were formulated, and Fresnel, using the Huygens principle, explained the rectilinearity of light and diffraction (Huygens-Fresnel principle). Based on the polarization of light established in the experiments, Young and Fresnel were the first to put forward the idea of the transverse nature of light waves. The works of Faraday and Maxwell finally proved the wave nature of light. Hertz's discovery of electromagnetic waves made wave electromagnetic theory indisputable. The electromagnetic wave theory of light was further

developed in the late 19th century in the works of the French physicist François-Pierre Le Roux (1862), the German physicists August Kundt, and W. Zemmeter (1872), the German physicists Paul Drude, Hermann von Helmholtz and the Dutch physicist Hendrik Lorentz (90s of 19th century). In 1899, Russian physicist Pyotr Lebedev discovered and explained in terms of electromagnetic theory the pressure of light. Finally, the wave electromagnetic theory of light served as the starting point for Einstein's theory of relativity (1905). This, in particular, was facilitated by the work that revealed the contradictions between Galileo's principle of relativity and Maxwell's electromagnetic theory. As early as 1804, Young showed that the wave theory of light requires the introduction of the idea of an absolutely stationary ether. Fresnel (1818), François Arago (1810) and Hippolyte Fizeau (1890s), on the contrary, proved that a number of optical phenomena (for example, independence of the index of refraction of light from the motion of bodies) requires the assumption of partial entrainment of the ether by moving bodies. The same conclusion was reached by Lorentz, who in 1896 created the theory of electrodynamics of moving media. Finally, the Michelson-Morley experiments showed that the idea of the ether requires either a rejection of the principle of relativity or of Maxwell's theory.

All of these contradictions were resolved in Einstein's theory of relativity. Meanwhile, just when it seemed that the electrodynamic theory of light had won its final victory, i.e., by the end of the 19th century and the beginning of the 20th century, the first signs of its crisis appeared.

In particular, it turned out that this theory in explaining the processes of emission and absorption of light contradicts the law of conservation of energy. Analyzing this phenomenon, Planck came to the conclusion in 1900 that the elementary vibrating system (atom, molecule) gives or receives wave energy from the electromagnetic field **not continuously**, but in portions, by photons proportional to the **frequency of oscillations**. Developing the ideas of Planck, Einstein in 1905 attributed to the photons, in addition to energy, properties of mass and momentum. He also showed that the photoelectric effect can be explained only on the basis of the notion that electrons, absorbing photons of light, change their motion and create an electric current of conduction

(photoelectric current). In other words, according to the Planck and Einstein hypotheses, the following relations take place

$$\left.\begin{aligned} \frac{E}{\omega} &= \hbar \\ P &= \frac{\hbar\omega}{c} \\ m &= \frac{\hbar\omega}{c^2} \end{aligned}\right\} \qquad (6.107)$$

It turned out that the total volumetric radiation density ρ, equal to the radiation energy per unit volume, and its proportional value of the emissivity of an black body, summed over all frequencies of the frequency range of radiation (from zero to infinity), are equal to

$$U = \sigma T^4; \qquad (6.108)$$

$$\sigma = \frac{2\pi^5 k^4}{15c^2h^3} \qquad (6.109)$$

where

U-emissivity (power of radiation of the unit of the black body area, W m^{-2});

σ is the Stefan–Boltzmann constant;

k is the Boltzmann constant (1.38x 10 - 23J K^{-1});

h is Planck's constant (6.58x 10^{-34} J s);

c is the speed of light in a vacuum (3x108 m s^{-1}). The law of radiation (6.108) is called the Stefan-Boltzmann law. It is in perfect alignment with the experiment. It made it possible to calculate the value of $\hbar$.

Since it follows from (6.107) that

$$\mathcal{E} = \hbar\omega = \frac{h\omega}{2\pi} \cong 6{,}38 \cdot 10^{-34} \cdot \omega \qquad (6.109)$$

Einstein, based on the concept of photons, formulated his famous law of photoelectric effect, according to which

$$\mathcal{E} = \hbar\omega - \mathcal{E}_i \qquad (6.110)$$

where

$\mathcal{E}$ is the initial energy of the photoelectron;

$\mathcal{E}_i$ the ionization energy of the atom.

Einstein's law resolved the so-called ultraviolet crisis. Theoretically, the photoelectric current should depend on the degree of illumination of the irradiated sample (the square of the wave amplitude).

In fact, when irradiating the sample with a monochromatic wave shifted to the red side of the spectrum of any intensity, there was no photoelectric current at all. However, when the wavelength shifted toward the ultraviolet part of the spectrum, starting from some well-defined wavelength value, photoelectric current occurred even in low light. In other words, it turned out that the photoelectric current increases with decreasing wavelength and is completely independent of the intensity. According to Bohr, the same law applies to the emission and absorption of energy by an atom. The hypotheses of Bohr (for the hydrogen atom), Einstein and Planck were fully confirmed by physical experiment and came to be regarded as scientifically proven fact.

6.6.4. De Broglie Waves

In 1912, the Danish physicist Niels Bohr, relying on the hypotheses of Planck and Einstein, created a model of the atom, according to which the atom emits light of appropriate wavelength X, when under the action of absorbed energy it goes from an excited state with energy *hv*, to a stationary state with energy conventionally assumed to be zero. In this case

$$\mathcal{E}_i = kv_i \qquad (6.111)$$

The magnitude of energy $h\nu_i$ corresponds to the energy of the photon, and the excitation levels form a discrete series (i = 1, 2, 3, ...). The energy corresponding to the (i ± 1) levels is at band gap for the atom.

In 1924, the French physicist Louis de Broglie, based on the hypotheses of Planck, Einstein, and Bohr, suggested that elementary particles (electrons, nucleons), just like light, must exhibit both corpuscular and wave properties. In this sense, their corpuscular properties and wave properties must be linked by the same equations that characterize photons. In other words, for elementary particles, as well as for photons, according to the de Broglie hypothesis, the following relations must take place

$$\omega = \frac{\mathcal{E}}{\hbar};\ \lambda = \frac{2\pi\hbar}{P} \tag{6.112}$$

Waves associated with free-moving elementary particles are called, as mentioned above, de Broglie waves.

For low-energy particles whose relative speed is much lower than the speed of light in a vacuum

$$\upsilon << c, \tag{6.113}$$

it is possible, according to (6.112), to write that

$$\lambda = \frac{2\pi\hbar}{m\upsilon} \tag{6.114}$$

In other words, the wavelength compared to the particles is inversely proportional to their mass. So, for example, the wavelength $\lambda_э$, which corresponds to the electron, with a high enough energy for it, (10 -10^4) eV, is, according to the relation (6.114)

$$\lambda_э = (4{,}0 - 0{,}1)\acute{A} \tag{6.115}$$

(Angstrom is a unit of length; $1A = 10^{-10}$. This length is in the spectrum of X-rays. As the mass of a particle increases, the wavelength of the wave it is matched to sharply decreases. Already for a hydrogen molecule it is equal to $\lambda_{H2=}(0{,}07 - 0.002)\acute{A}$. A small macroparticle with

a mass of 1 g at a paltry energy equal to the energy of an electron, the wavelength corresponds to k $\lambda \approx 10^{-15} \acute{A}$

Waves of such short length are in principle unobservable even with instruments of any high sensitivity and resolution. For this reason, wave properties of macroparticles (macrobodies) cannot be detected.

De Broglie's hypothesis found brilliant experimental confirmation. Already in 1927, in experiments by American physicists Clinton Davisson and Lester Germer, it was found that a beam of electrons with an energy of 100 - 150 eV with $\lambda \approx (1 - 1{,}5)\acute{A}$ diffracts on nickel crystals, playing the role of a diffraction lattice. Experiments on interference and diffraction of electron beams, as well as beams of other elementary particles, were repeated several times and always led to the same results, which unambiguously confirmed the wave properties of particles.

At present, wave properties of elementary particles are widely used in practice, for example, in electron microscopes, as well as in devices designed to study the structure of matter. Let us consider in more detail one of many experiments on the interference of light and electron beams using a two-slit Fresnel interferometer (see Section 6.5.3). In a simplified version, this interferometer (Fig. 6.24) is a plate A opaque to light with two close small holes (slits) 1 and 2, to which a beam of monochromatic light or electrons from the source S is directed. An interference pattern appears on the screen (indicator) B, installed to the right of the plate A.

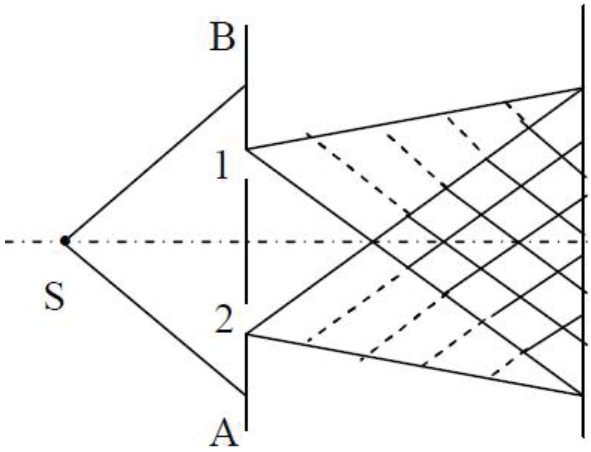


Figure 6.24.

According to Fresnel, the wave falling on plate A creates secondary coherent waves in holes 1 and 2, which continuously add up and interfere.

Let the secondary waves coming from holes 1 and 2 be described by the wave functions

$$\left.\begin{aligned}\psi_1 &= A_1 cos(\omega t \pm kx_1)\\ \psi_2 &= A_2 cos(\omega t \pm kx_2)\end{aligned}\right\} \tag{6.116}$$

In complex form

$$\left.\begin{aligned}\psi_1 &= A_1 e^{j(\omega t \pm \varphi_1)}\\ \psi_2 &= A_2 e^{j(\omega t \pm \varphi_1)}\end{aligned}\right\} \tag{6.117}$$

where

$$\varphi_1 = kx_1;\ \varphi_2 = kx_2$$

or

$$\left.\begin{aligned}\psi_1 &= \hat{A}_1 e^{j\omega t}\\ \psi_2 &= \hat{A}_2 e^{j\omega t}\end{aligned}\right\} \tag{6.118}$$

where

$$\left.\begin{aligned}\hat{A}_1 &= \hat{A}_1 e^{jkx_1}\\ \hat{A}_2 &= \hat{A}_2 e^{jkx_2}\end{aligned}\right\} \tag{6.119}$$

At interference, waves ψ_1 and ψ_2 add up so that the resulting wave

$$\psi = \psi_1 + \psi_2.$$

In complex form

$$\psi = \hat{A}_1 e^{j\omega t} + \hat{A}_2 e^{j\omega t} = \left(\hat{A}_1 + \hat{A}_2\right) e^{j\omega t} \quad (6.120)$$

In other words, the resulting amplitude

$$\hat{A} = \hat{A}_1 + \hat{A}_2 \quad (6.121)$$

It was said above that the intensity is determined by the square of the amplitude

$$J_1 = \left|\hat{A}_1\right|^2;\ J_2 = \left|\hat{A}_2\right|^2;\ J_{12} = \left|\hat{A}_1 + \hat{A}_2\right|^2 \quad (6.122)$$

where

J_1 is the intensity of the wave that only passed through hole 1;

J_2 is the intensity of the wave that only passed through hole 2;

J_{12} is the intensity of the wave that passed simultaneously through holes 1 and 2 and gives rise to two secondary coherent waves, which interfere when added together.

It follows from relation (6.122) that the intensity of each individual wave $J_i = |A_i|^2 = A_i A*_i$ ($A*_i$ is the amplitude complex-conjugate to amplitude A_i) is independent of phase and is determined by a number expressed as the square of the actual amplitude. The intensity of the total wave resulting from the interference of secondary coherent waves is determined not only by the sum of the squares of the amplitudes, but also by the phase difference of the interfering waves, i.e.

$$J_{12} = \left|\hat{A}_1 + \hat{A}_2\right|^2 = A_1^2 + A_2^2 + 2A_1A_2cos(\varphi_2 - \varphi_1) \quad (6.123)$$

The quantity $2A_1A_2cos(\varphi_2 - \varphi_1)$ is called the interference term. The electron beam behaves similarly in the interference experiment. This means that the same electron at interference simultaneously passes through holes 1 and 2, which for a corpuscle is impossible in principle.

On the other hand, it turns out that only a multiple number of electrons can pass through hole 1 (with hole 2 closed) and vice versa. This means that in the absence of interference, the electrons behave like ordinary corpuscles.

The considered interference experience allows us to introduce a new function ψ (r, t) of the type (6.116) or (6.117) to characterize the state of a microparticle. In the one-dimensional version this function has the form

$$\psi(x,t) = ReCe^{-j(\omega t - kx)} \tag{6.124}$$

where k is the wave number.

According to relation (6.113)

$$k = \frac{P}{\hbar} = \frac{2\pi}{\lambda} \tag{6.125}$$

where

λ is the wavelength matched to the microparticle;

π is the impulse of the microparticle

It follows from this that the behavior of a microparticle, unlike a pure corpuscle, is nondeterministic. This manifests itself in the fact that the same particles behave differently under the same conditions for no reason whatsoever. In the above experiment with beams of perfectly identical electrons, some particles pass through slit 1, while others pass through slit 2.

It is impossible in principle to predict in advance how a given particle will behave. Its behavior is truly random. To predict it, it is necessary to apply a probabilistic approach. The German physicist Max Born, considering experiments on interference of microparticles, for the first time proposed a way to calculate the specified probability. In so doing, he reasoned as follows.

Let an elementary volume *dV* spanning some point r(x,y,z) be given. Let *ip*(r, t) be the wave function describing the state of the microparticle inside this volume. Since the quantity $|\psi(r,t)|^2$, according

to (6.122) determines the intensity of the wave, which, according to the de Broglie hypothesis, describes the state of a microparticle, the probability of finding a microparticle at time t inside the volume dV is

$$d\rho = |\psi(r,t)|^2 dV, \qquad (6.126)$$

The probability of detecting a microparticle at a given point r(x, y, z) is defined from (6.126) as the value of its density

$$\frac{d\rho}{dV} = |\psi(r,t)|^2 \qquad (6.127)$$

Thus, the square of the amplitude of the wave function takes on the meaning of the probability of detecting a microparticle at a particular point of the volume in which this function is given. On the other hand, it is obvious that a microparticle at time t is reliably detected in at least one point of volume V. It follows that the probability of detecting a particle at any point in the volume V, is equal to 1. This probability is equal to the sum of the probabilities of detecting a particle in each of the points of the volume V.

Thus, it follows from (6.124) that

$$\rho = \int_V |\psi(r,t)|^2 dV = 1 \qquad (6.128)$$

The relation (6.128) is called the **normalization rule** for the wave function of microparticles.

With this in mind, the wave function describing the state of the microparticle is also called the **probability amplitude**.

On the other side,

$$\psi = \psi_1 + \psi_2 ,$$

where ψ_1, ψ_2, ψ are, respectively, the amplitudes of the probabilities of a particle passing through hole 1, hole 2, and both holes simultaneously.

Summarizing the result obtained, it can be stated that if under given conditions a physical system can be in different states

corresponding to wave functions ψ_1, ψ_2, ψ_3,...ψ_k,..., then under the same conditions states with a wave function are also possible

$$\psi = \sum_k c_k \psi_k \tag{6.129}$$

where c_k are some complex numbers.

6.6.4.1. Time-independent Schrödinger equation

De Broglie waves describe the state of a particle only in the case of its free motion. Accordingly, the equation that the de Broglie wave satisfies is also compared to the free motion of a microparticle with constant energy g and momentum

$$P = \sqrt{2m\mathcal{E}}$$

It follows from (6.24) that the wave function describing a de Broglie wave in the one-dimensional version in the complex form is

$$\psi(x,t) = Ae^{j\left(\omega t + \frac{P}{\hbar}x\right)}$$

or

$$\psi(x,t) = Ae^{j\omega t}e^{j\frac{P}{\hbar}x} = \hat{A}e^{j\frac{P}{\hbar}x}$$

In the case of fixed time, when $\hat{A} = const$, the wave function $\psi(x)$ satisfies the wave equation, which is obtained from the equation of free oscillations (see Section 5.1.2) by replacing t by ω and o by k. Since, according to (6.125)

$$k = \frac{P}{\hbar}$$

then the one-dimensional wave equation for the function y/(x) describing the de Broglie wave at a given time takes the form

$$\frac{d^2\psi(x)}{dx^2} + \frac{P^2}{\hbar^2}\psi(x) = 0 \tag{6.130}$$

or

$$\frac{d^2\psi}{dx^2} + \frac{2m\mathcal{E}}{\hbar^2}\psi = 0 \qquad (6.131)$$

Let us consider a microparticle moving in a stationary external field U(x) with constant energy Ɛ. For a given particle according to the law of conservation of energy

$$p = \sqrt{2m[\mathcal{E} - U(x)]}$$

The wave equation in this case takes the form

$$\frac{d^2\psi}{dx^2} + \frac{2m[\mathcal{E} - U(x)]}{\hbar^2}\psi = 0 \qquad (6.132)$$

Equation (6.132) is written for a fixed time and is therefore called the **stationary** Schrödinger equation in honor of the Austrian physicist Erwin Schrödinger, who first proposed and studied it.

In the general case of three-dimensional particle motion, the time-independent Schrödinger equation takes the form

$$-\frac{\hbar^2}{2m}\Delta\psi + U(x,y,z)\psi = \mathcal{E}\psi \qquad (6.133)$$

where $\Delta\psi$ is the Laplace operator

The time-independent Schrödinger equation describes the behavior of a microparticle, at a given moment in time, in a stationary external field $U(\vec{r})$.This state is determined by the type of external field and the value of the wave function at its boundaries. Let us consider the stationary states of a particle for a number of very important types of one-dimensional fields of a given configuration.

1. Potential wall

A one-dimensional potential wall is an idealized potential field U(x) that satisfies the following conditions

$$U(x) = \begin{cases} 0 & x < 0 \\ U_0 & 0 \le x < \infty \end{cases} \qquad (6.134)$$

The potential wall is shown in Fig. 6.25. Let us first consider the case where the energy of the particle $\mathcal{E} = \mathcal{E}_1 > U_0$

A potential wall generally forms a barrier for a free-moving microparticle. The case in question is called **over-barrier passage**. Let's assume that the particle moves from left to right. In the region of x<0, the value of U(x) = 0 and the one-dimensional Schrödinger equation has the following form

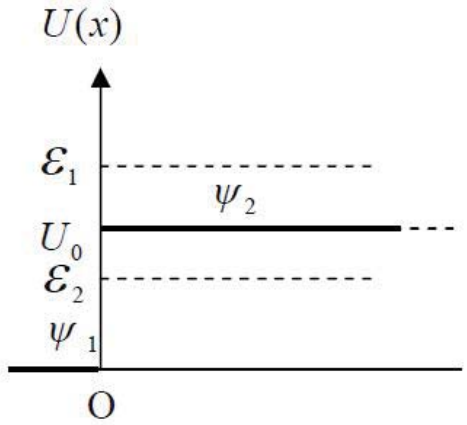


Figure 6.25.

$$\frac{d^2\psi(x)}{dx^2} + \frac{2m\mathcal{E}}{\hbar^2}\psi(x) = 0 \qquad (6.135)$$

where

$$\frac{2m\mathcal{E}}{\hbar^2} = k^2 \qquad (6.136)$$

The general solution of the resulting equation, as we know, can be written as follows

$$\psi(x) = \psi_1 + \psi_2 = C_1 e^{jkx} + C_2 e^{-jkx} \qquad (6.137)$$

The first term of this equation describes the wave incident on the wall, and the second term of the equation, as is easy to see, describes the reflected wave.

Thus, even though the particle moves over the barrier, it is partially reflected from the wall.

Obviously, this behavior of the particle is a consequence of its wave properties. For a corpuscle this is impossible in principle.

In the region x ≥ 0, the Schrödinger equation is modified and written in the form

$$\frac{d^2\psi(x)}{dx^2} + \frac{2m(\mathcal{E}_1 - U_0)}{\hbar^2}\psi = 0 \tag{6.138}$$

where

$$\frac{2m(\mathcal{E}_1 - U_0)}{\hbar^2} = k_1^2 \tag{6.139}$$

The solution of equation (6.139) will be, as in the case (6.136),

$$\psi(x) = C_1' e^{jk_1 x} + C_2' e^{-jk_1 x} \tag{6.140}$$

Since, however, there can only be a passing wave in the region x ≥ 0, then $C_2' = 0$ and

$$\psi(x) = C_1' e^{jk_1 1x} \tag{6.141}$$

* It follows from physical considerations that the wave function and its derivatives are continuous everywhere.

Substituting x = 0 into the continuity condition gives

$$\left.\begin{aligned} C_1 + C_2 &= C_1' \\ k(C_1 - C_2) &= k_1 C_1' \end{aligned}\right\} \tag{6.142}$$

Solution of equation (6.142), taking into account that the reflection coefficient

$$R = \frac{|C_2|^2}{|C_1|^2}$$

gives

$$R = \frac{|C_2|^2}{|C_1|^2} = \frac{(k - k_1)^2}{(k + k_1)^2} \tag{6.143}$$

Assuming that the transmission coefficient is equal to

$$D = 1 - R$$

We find that

$$D = 1 - R = \frac{4k_1k_2}{(k + k_1)^2} \tag{6.144}$$

In the case of

$$\mathcal{E} = \mathcal{E}_2 < U_0 \tag{6.145}$$

then

$$k_1^2 < 0$$

This means that the value of k_1 is an imaginary number. In this case, the wave function for the traveling wave at $x \geq 0$

$$\psi(x) = C_1' e^{-|k_1|x} \tag{6.146}$$

In other words, the probability of detecting a particle whose energy is less than the height of the wall barrier on the right side of the wall is not zero, but quickly fades so that it can in principle only be detected near the wall (Fig. 6.26)

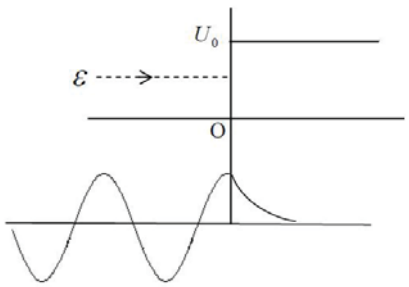
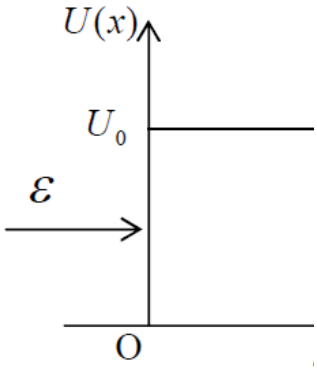


Figure 6.26.

2. Rectangular potential barrier

A rectangular potential barrier (Fig. 6.27) is an idealized one-dimensional potential field that satisfies the following conditions



$$U(x) = \begin{cases} 0 & x < 0 \\ U_0 & 0 \le x \le 1 \\ 0 & x > 1 \end{cases} \qquad (6.147)$$

Figure 6.27.

Let a microparticle with energy $0 < \mathcal{E} < U_0$ move from left to right.

Outside the potential barrier, in the regions *x < 0 and x > l*, the time-independent Schrödinger equation coincides with equation (6.135).

At the same time, if

$$x < 0,$$

then

$$\psi(x) = C_1 e^{jkx} + C_2 e^{-jkx} \qquad (6.148)$$

and if

$$x > 1,$$

then

$$\psi(x) = C_3 e^{jkx} + C_4 e^{-jkx} \qquad (6.149)$$

where

$$k = \frac{\sqrt{2m\mathcal{E}}}{\hbar}$$

Inside the barrier, in the region *0 < x < 1*, the Schrödinger equation coincides with equation (6.138), and the wave function

$$\psi(x) = C_5 e^{k_1 x} + C_6 e^{-k_1 x} \tag{6.150}$$

where

$$k_1 = \frac{\sqrt{2m(U_0 - \mathcal{E})}}{\hbar} \tag{6.151}$$

The wave outside the barrier, in the region $x > l$, is only a passing wave and corresponds to a particle that moves in this region from left to right, so here $C_4 = 0$, and

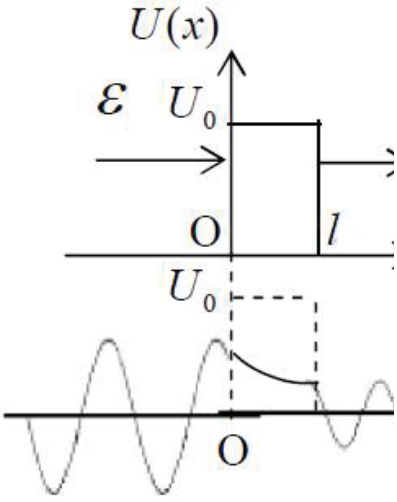
$$\psi(x > l) = C_3 e^{jkx}$$

In the region of x < 0

$$\psi(x < 0) = C_1 e^{jkx} + C_2 e^{-jkx}$$

In other words, a particle hitting a barrier corresponds to two waves, the incident and reflected ones. As in the case of a potential wall, the incident wave penetrates the region of the potential barrier, $0 \le x \le l$, in which it becomes damped. However, after reaching the wall x = l, the wave passes beyond the potential barrier and penetrates into the region x > l, where $k^2 > 0$. Therefore, in the region x > 1, it again acquires its periodicity and then propagates as a undamped traveling wave with lower intensity, as shown in Figure 6.28.

From the condition of continuity



$$\left.\begin{aligned} C_1 + C_2 &= C_3 + C_4 \\ k(C_1 - C_2) &= k(C_3 - C_4) \\ C_3 e^{kl} + C_4 e^{-kl} &= C_5 e^{jkl} \\ k_1 C_3 e^{kl} - k_1 C_4 e^{-kl} &= jk C_5 e^{jkl} \end{aligned}\right\} \tag{6.152}$$

Figure 6.28.

* Solving the system (6.152), we obtain that

1) reflection coefficient in the region x < 0

$$R = |b_1|^2 = \frac{|C_1|^2}{|C_1|}$$

2) probability of particle permeation through the barrier

$$D = |a_3|^2 = \frac{|C_5|^2}{|C_1|}$$

at the same time

$$a^3 = -j\frac{4k_1 k e^{jkl}}{(k_1 - jk)^2} e^{-k_1 l}$$

and

$$D = -j\frac{16k_1^2 k^2}{(k_1^2 - k^2)^2} e^{-2kl} \qquad (6.153)$$

$$D \approx e^{-2kl} \qquad (6.154)$$

It follows from (6.153) that the degree of particle penetration through the barrier decreases exponentially with increasing barrier width. The phenomenon of a particle leaking through a potential barrier is called the **quantum tunnelling**.

3. Potential well

Let *U(x) = 0* in the bounded space from *x = 0 to x = l* (Fig. 6.29a). If the energy of a particle moving in the positive side of the Ox axis over the potential well is $\mathcal{E} > 0$, then the probability of its reflection from the walls of the potential well, as was found above, is not zero. The energy of the particle above the well has a continuous emission spectrum everywhere.

The particle behaves differently inside the potential well, where its energy $\mathcal{E} < U(x)$

It was found above that outside the well, the sides of which can be regarded as potential walls, the wave function should damp down. To

determine the damping conditions, let us find the solution of the Schrödinger equation outside and inside the potential well. To simplify the problem, we consider an idealized one-dimensional potential well of infinite depth (Fig. 6.29b).

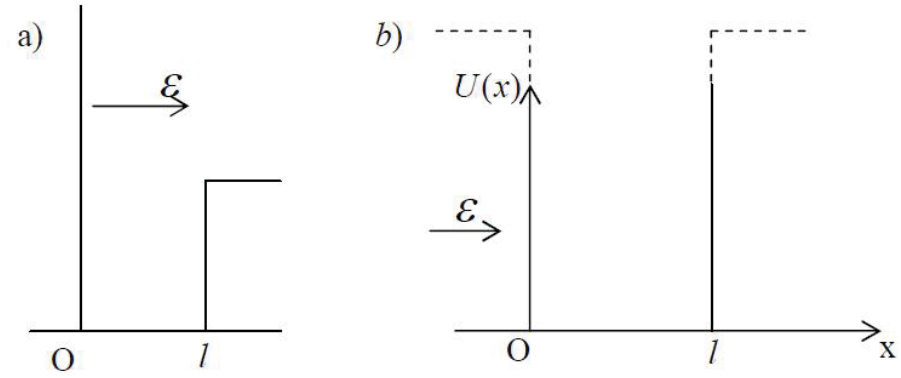


Figure 6.29.

As can be seen from the figure, the boundary conditions for such a well take the form

$$U(x) \to \begin{cases} +\infty & x < 0 \\ 0 & 0 \le x \le l \\ +\infty & x > l \end{cases} \tag{6.155}$$

Let us find a solution of the Schrödinger equation for each of these three regions, assuming that $\mathcal{E} > 0$.

In the region of x < 0

$$\frac{d^2\psi}{dx^2} - a^2\psi = 0 \tag{6.156}$$

where

$$a^2 = \frac{2m}{\hbar^2}(U - \mathcal{E}) > 0 \tag{6.157}$$

The solution to equation (6.156) is

$$\psi = C_1 e^{kx} + C_2 e^{-kx} \tag{6.158}$$

Since x < 0 and k > 0, then at $x \to -\infty$

$$e^{kx} \to 0, and\ e^{-kx} \to \infty \tag{6.159}$$

On the other hand, the wave function must be finite everywhere. This requires that $C_2 = 0$. Hence, taking into account (6.159), we obtain that with fast damping,

$$\psi(x < 0) \to 0 \tag{6.160}$$

We obtain the same result obviously for the region x > l

$$\psi(x > l) \to 0 \tag{6.161}$$

For the region

$$0 \leq x \leq l$$

takes place

$$\frac{d^2\psi}{dx^2} + k^2\psi = 0 \tag{6.162}$$

Where

$$k^2 = \frac{2m\mathcal{E}}{\hbar^2} > 0 \tag{6.163}$$

The solution of equation (6.162) is thus a harmonic function (see Section 5.1)

$$\psi(x) = A\sin(kx + \varphi) \tag{6.164}$$

Since this function is zero at the boundaries *x = 0 and x = l*, it is necessary and sufficient that

$$\left.\begin{aligned} A\sin\varphi = 0 \\ A\sin(kl + \varphi) = 0 \end{aligned}\right\} \tag{6.165}$$

From the first equation it follows that *φ = 0*, so

$$A\sin kl = 0 \tag{6.166}$$

Since A ≠ 0 (otherwise $\psi(x) \equiv 0$ in the entire region from $+\infty$ до $-\infty$, which would indicate the absence of the microparticle in the potential well), then

$$\sin kl = 0 \tag{6.167}$$

From which

$$k = \frac{\pi}{l} n, where\ n = 0,1,2,3, \ldots \tag{6.168}$$

Substituting (6.168) into (6.163) gives

$$\frac{2m}{\hbar^2} \mathcal{E}_n = \frac{\pi^2}{l^2} \cdot n^2 \tag{6.169}$$

or

$$\mathcal{E}_n = \frac{\pi^2 \hbar^2}{2ml^2} \cdot n^2 \tag{6.170}$$

Thus, when a particle is in a potential well, its energy takes on only discrete values $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \ldots$ The natural number *n = 1, 2, 3, ...* is called the **quantum number**. The quantum number determines the energy state or allowed energy level of a particle inside a well. The result is called **quantization**. Quantization thus means that the energy spectrum of a particle inside a potential well is discrete.

To determine the amplitude A of the wave function *ψ(x)*, we use the rule of its normalization (6.128), according to which (in the one-dimensional case)

$$\int_{-\infty}^{+\infty} |\psi(x)|^2 \, dx = 1 \tag{6.171}$$

Since outside the pit $\psi(x) = 0$, (6.171) takes the form i

$$\int_0^l |\psi(x)|^2\, dx = 1 \qquad (6.172)$$

Substitution $\psi(x)$ from (6.164) into (6.172) taking into account (6.165) and (6.155) gives

$$\int_0^l A^2 \sin^2 \frac{\pi n}{l} x\, dx = 1 \qquad (6.173)$$

from which

$$A = \sqrt{\frac{2}{l}} \qquad (6.174)$$

At $0 \le x \le l$

$$\psi_n(x) = \sqrt{\frac{2}{l}} \sin\left(\pm \frac{\pi n}{l} x\right) \qquad (6.175)$$

where $n = 1,2,3,...$

Outside the potential well, the wave function is zero. The nominal value of the energy state at $n = 1$ is the minimum possible. This state is called the **basic state**. When $n = 2, 3, ...$ the state is called **excited**. It also follows from (6.175) that for $\mathcal{E} < 0$ the problem has no solution. It can be shown that the results obtained are true not only for one-dimensional, but also for three-dimensional barriers and potential wells.

An example of the state of a microparticle of matter inside a potential well is an electron bound inside an atom. Such an electron is known to be under the action of the Coulomb attraction of the nucleus, whose potential energy forms a potential well, and the energy of the electron is quantized and forms a discrete spectrum of energy levels.

6.6.4.2. Time-dependent Schrödinger equation

So far we have considered only possible quantum states of particles with constant energy Ɛ, which derive from the solution of the time-independent Schrödinger equation and set its state at a given time. In dynamics, the state, including the energy of a particle Ɛ, changes with time. In the general case, see (6.124)

$$\psi(r,t) = Ae^{-j\omega t}e^{jkr}$$

Since

$$\psi(x,t) \sim e^{-j\omega t}e^{jkx}$$

then

$$\frac{\partial \psi}{\partial t} = -j\omega\psi$$

Let us multiply the last equation by $\hbar$, then

$$\hbar\frac{\partial \psi}{\partial t} = -j\omega\hbar\psi$$

or

$$\text{Ɛ}\psi = j\hbar\frac{\partial \psi}{\partial t}$$

Substitution in (6.133) gives

$$j\hbar\frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\omega + U\psi \qquad (6.176)$$

Equation (6.176) is called the **time-dependent Schrödinger equation**. To describe the behavior of microparticles, the **time-dependent Schrödinger equation** plays the same role as the equation of motion in the mechanics of macro-processes.

Equation (6.176) is often written as

$$j\hbar\frac{\partial \psi(r,t)}{\partial t} = -\hat{H}\psi \qquad (6.177)$$

where $\hat{H}$ is the Hamiltonian operator.

$$\widehat{H} = -\frac{\hbar^2}{2m}\Delta + U \qquad (6.178)$$

$\widehat{H}$ is an analogue of the Hamiltonian introduced in classical mechanics for the non-Newtonian notation of the equation of motion (see Section 2.7.7).

6.6.5. Acoustic waves

Acoustics is the field of physics that studies elastic vibrations and waves in the frequency range from zero to 10^{11} to 10^{13} Hz. Acoustic waves also include sound, which covers the spectrum of elastic waves between 20 and 20,000 Hz.

Acoustics originated in ancient times as the study of sound. The physical properties of sound were studied by Pythagoras (6th century BC) and Aristotle (4th century BC). In the Middle Ages, sound waves were explored by Leonardo da Vinci (15th-16th centuries) and Galileo (17th century).

The general theory of radiation and propagation of sound waves in a medium appeared at the end of the 17th century. Huygens, Hooke, Galileo, the French scientist Marin Mersenne, and many others played a major role in its creation.

Acoustic waves and, in particular, sound refer to elastic waves propagating in solid, liquid and gaseous media.

When a solid elastic medium is deformed, mechanical waves arise in it, which transfer the deformation from one point to another. In this case there is also a transfer of elastic strain energy.

In liquids and gases, which have volume elasticity but no shape elasticity (see sections 2.9.2, 2.9.3), only longitudinal waves of tension and compression can propagate. An example would be sound waves.

Transverse shear waves also propagate in solids. At the boundaries of media, surface waves, which are a combination of longitudinal and transverse waves, can also occur.

6.6.5.1. Acoustic wave equations

In the general case acoustic waves propagate in a solid material medium, which is characterized by three fields - two scalar, pressure and density,

$$p = p(\mathrm{x}, y, z, t), \qquad \rho = \rho(\mathrm{x}, y, z, t)$$

and one vector field of velocities

$$\vec{v} = \vec{v}(\mathrm{x}, y, z, t).$$

The connection between these fields is considered in the section "Continuum mechanics" and in the equations of aerohydrodynamics (see Section 2.9.6).

In the following, using these equations, we will neglect the terms that take into account the effects of the gravitational field and the viscosity of the medium. In this simplified form, these equations can be written as follows

* 1) equations of motion

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v}, \nabla)\vec{v} = -\frac{\nabla p}{\rho} \tag{6.179}$$

(see Equation 2.233).

2) the equation of continuity

$$\frac{\partial \rho}{\partial t} + div(\rho \vec{v}) = 0 \tag{6.180}$$

Let us denote the values of pressure and density of the unperturbed medium that existed before the arrival of the wave by p_0 и ρ_0, and the changed values of pressure and density by p и ρ, such that

$$\mathrm{p}' = \mathrm{p} - \mathrm{p}_0 \text{ и } \rho' = \rho - \rho_0.$$

We consider wave processes resulting from small perturbations of the medium, at which

$$\rho \ll \rho_0; \quad \rho' \ll \rho_0$$

In this case, the perturbations do not violate the homogeneity and isotropy of the medium, and the hydrodynamic equations above are simplified and take the form

$$\left.\begin{aligned} \frac{\partial \upsilon}{\partial t} + \frac{1}{\rho_0}\nabla p' = 0 \\ \frac{\partial \rho'}{\partial t} + \rho_0 div\vec{v} = 0 \end{aligned}\right\} \tag{6.181}$$

The joint solution of the latter equations gives

$$\left.\begin{aligned} \nabla p' - \frac{1}{c^2}\frac{\partial^2 p'}{\partial t^2} = 0 \\ \nabla \rho' - \frac{1}{c^2}\frac{\partial^2 \rho'}{\partial t^2} = 0 \\ \nabla \upsilon' - \frac{1}{c^2}\frac{\partial^2 \upsilon}{\partial t^2} = 0 \end{aligned}\right\} \tag{6.182}$$

(6.182) are wave equations that describe wave propagation in a medium with velocity *c* (not to be confused with the speed of light in a vacuum c, which is a world constant).

Thus, we conclude that small perturbations of pressure, density and velocity created in some region of a continuous material medium propagate in it in the form of acoustic waves.

In addition, it can be shown (we accept without proof) that

$$p' = c^2\rho' \tag{6.183}$$

In the general case, the wave equation (for any of the wave characteristic quantities) can be written as

$$\Delta\psi - \frac{1}{c^2}\frac{\partial^2 \psi}{\partial t^2} = 0 \tag{6.184}$$

The solution of this equation in the one-dimensional version is the wave function

$$\psi(x,t) = \psi(x - ct)$$

Substituting this solution into the initial hydrodynamic equations for small perturbations gives

$$\upsilon = \frac{p'}{Z}$$

where Z is the value of the so-called acoustic impedance

$$Z = \rho_0 c$$

The wave function is harmonic and in the general case has the form

1) for the one-dimensional case

$$\psi(x,t) = Re\left|Ae^{j(kx-\omega t)}\right| \qquad (6.185)$$

2) for the three-dimensional case

$$\psi(\vec{r},t) = Re\left|Ae^{j[(kx-\omega t)]}\right|$$

$$\vec{k} = \frac{2\pi}{\lambda}\vec{n} \qquad (6.186)$$

where

λ is the wavelength;

ω is the cyclic frequency;

$\vec{n}$ is the unit vector of the wave propagation direction. *

6.6.5.2. Characteristics of acoustic waves

The speed of light in a vacuum, as we have seen, is a constant. The speed of propagation of acoustic, including sound waves, is determined by the characteristics of the medium, its elastic properties. Here are semi-empirical formulas for calculating the speed of acoustic waves and, in particular, sound, in different media, without proof. The speed of acoustic wave propagation:

- in gaseous media

$$c = \sqrt{\frac{\gamma p_0}{\rho}} \tag{6.187}$$

where γ is the relative displacement.

- in liquids

$$c = \sqrt{\frac{nA}{\rho_0}} \tag{6.188}$$

where *A and n* are constants for different liquids;

- in solids

$$c = \sqrt{\frac{E}{\rho_0}} \tag{6.189}$$

where *E* is the modulus of elasticity of longitudinal compression.

Speed of sound propagation:

- in liquids and gases

$$c = \sqrt{\frac{K}{\rho}} \tag{6.190}$$

where *K* is the bulk modulus of elasticity.

In ideal gas

$$c = \sqrt{\frac{\eta RT}{\mu}} \tag{6.191}$$

where

R is the universal gas constant;

η is the adiabatic index;

μ is the molar mass.

In the fields of acoustic waves, the elements of the medium are in motion and therefore have kinetic energy. In doing so, they undergo tensile and compressive strains. The energy density of the acoustic wave w is equal to the sum of the kinetic energy density in the initial state

$$w_k = \rho_0 \frac{\upsilon^2}{2}$$

and changes in the density of excess internal energy w'.

In other words,

$$w = w_k + w'.$$

* The density of excess internal energy is equal to the density of the work of the forces of excess pressure with a minus sign

$$dw' = -\frac{dA}{V} = -\frac{p'dV}{V} = \frac{c^2\rho' dVdm}{Vdm} = \frac{c^2\rho'}{\rho_0} d\rho \qquad (6.192)$$

where

dA is an element of work;

V is the volume.

By integrating (6.192), we obtain that

$$w' = \frac{c^2\rho'^2}{2\rho_0}$$

Total energy density in an acoustic wave

$$w = \frac{\rho_0 \upsilon^2}{2} + \frac{c^2\rho'^2}{2\rho_0} \qquad (6.193)$$

For acoustic waves, as well as for electromagnetic waves, the concept of Poynting vector $\vec{\Pi}$ is introduced by the formula

$$d\vec{\Pi} = \left[\vec{J}_w, d\vec{S}\right] \qquad (6.194)$$

where $\vec{J}_w$ is the vector of energy flux density

$$\vec{J}_w = \frac{d\mathcal{E}}{dSdt}\vec{e}_0; \tag{6.195}$$

$\vec{e}_0$ is the unit vector of the direction of energy flow.

From (6.191) and (6.195) it follows that

$$d\vec{\Pi} = \left[\frac{d\mathcal{E}}{dSdt}\vec{e}_0 dS \cdot \vec{n}_0\right] \tag{6.196}$$

from which

$$d\Pi = -\frac{d\mathcal{E}}{dt} \tag{6.197}$$

On the other hand, substitution (6.195) gives

$$\mathcal{E} = \int_v w dV = \frac{\rho}{2}\int_v v^2 dV + \frac{v^2}{2\rho_0}\int \rho'^2\, dV$$

$$\frac{d\mathcal{E}}{dt} = \frac{\rho_0}{2}\int 2\left(\vec{v}, \frac{d\vec{v}}{dt}\right) dV + \frac{v^2}{2\rho_0}\int 2\rho' \frac{dp'}{dt} dV$$

It follows from the equations of hydrodynamics that since

$$\frac{dv}{dt} = -\nabla p, and \;\; \frac{dp}{dt}\nabla v,$$

$$\frac{d\mathcal{E}}{dt} = -\int_v (\vec{v}, \nabla p')\, dV$$

$$-v^2 \int_v \rho'(\nabla, \vec{v}) dV$$

$$= -\int_v [(\nabla\vec{v}, \nabla p') + p'\nabla\vec{v}] dV$$

$$= -\int_V (\nabla, \vec{v}) dV = -\oint (p', \vec{v}, dS)$$

(see Field Theory, Appendix 3).

$$\Pi = \oint_S \left(\vec{J}_w, d\vec{S}\right) = \oint_S \left(p'\vec{v}, d\vec{S}\right)$$

or

$$\oint_S \left((-p'\vec{v} + \vec{J}_w\right), d\vec{S}) = 0,$$

from which

$$\vec{J}_w = p'\vec{v}. \tag{6.198}$$

The acoustic impedance of the medium Z, introduced above, is calculated by the formula

$$Z = \frac{p'}{v} \tag{6.199}$$

Substituting this expression into formula (6.198) gives the energy flux density

$$J_w = \frac{p'^2}{Z} \tag{6.200}$$

For periodic waves, the energy flux density Jw averaged over a period is

$$J_{av} = \frac{1}{Z}\frac{1}{T}\int_0^T p'^2 dt$$

Since the value of pressure P' satisfies the wave equation, then

$$p' = p'_m \sin\frac{2\pi}{T}t$$

Substitution yields

$$J_{av} = \frac{1}{2Z}p'^2_m \tag{6.201}$$

J_{av} is called the intensity of the acoustic wave. *

Acoustic waves obey the well-known laws of reflection and refraction.

Let monochromatic acoustic wave 1 (Fig. 6.30) fall perpendicularly to the interface between two media with coordinate x = 0. Let's denote the acoustic impedances of the media by Z_1 *and* Z_2

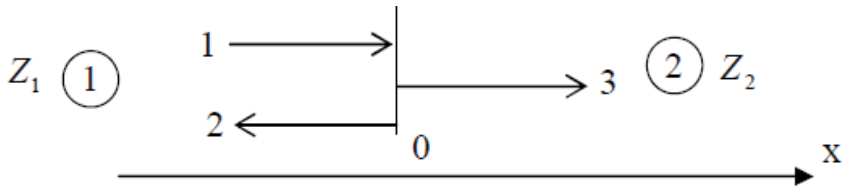


Figure 6.30.

This produces reflected 2 and refracted 3 waves.

Let it further

$$p'_1 = A_1 e^{j(\omega t - k_1 x)}$$

then

$$v_1 = \frac{A_1}{Z_1} e^{j(\omega t - k_1 x)} \tag{6.202}$$

Accordingly

$$p'_2 = A_2 e^{j(\omega t + k_1 x)};\ v_2 = \frac{A_2}{Z_1} e^{j(\omega t + k_1 x)}$$

$$p'_3 = A_3 e^{j(\omega t - k_2 x)};\ v_3 = \frac{A_3}{Z_2} e^{j(\omega t - k_2 x)}$$

Let the forces act on the interface from the wave side, the modules of which we denote by f_{12} *and* f_{21}. According to the third law of motion

$$f_{12} = f_{21}.$$

Let us define an element dS. In medium 1 the pressure is created by waves 1 and 2, and in medium 2 - only by wave 3, so

$$f_{21} = (p'_1 + p'_2)S \quad \text{in medium 1}$$

$$f_{12} = p'_3 S \quad \text{in medium 2}$$

it follows that

$$p'_1 + p'_2 = p'_3.$$

Accordingly

$$A_1 + A_2 = A_3.$$

It also follows from the continuity condition that

$$v_1 + v_2 = v_3$$

Let us denote $Z1 / Z2 = \alpha$, then after substitution we obtain

$$A_2 = \frac{\alpha - 1}{\alpha + 1} A_1 \tag{6.203}$$

$$A_3 = \frac{2\alpha}{\alpha + 1} A_1 \tag{6.204}$$

The energy coefficients of reflection R and refraction S can be written in the form

$$R = \frac{J_{2av}}{J_{1av}}; S = \frac{J_{3av}}{J_{1av}} \quad (6.205)$$

In the case when $\alpha = 1$, т. е. $Z_1 = Z_2$ we obtain that

$$R = 0 \text{ and } S = 1.$$

If $\alpha \to 0$ или $\alpha \to \infty$ then the resistances of the media are very different. In the first and second cases in this

$$R \to 1 \text{ and } S \to 0$$

Thus, at equal acoustic impedances, no reflected wave occurs, and all the energy of the incident wave is converted into the energy of the refracted wave.

On the contrary, when the acoustic impedance of the second medium is much greater than the first, the wave is completely reflected.

Then at $\alpha > 1$ $R > 0$ и $S = 0$ i.e. the reflection is in-phase, and at $\alpha < 1$ $r > 0$, но $S < 0$ i.e., the reflection is antiphase.

6.6.5.3. Elements of physiological acoustics

Subjectively, the frequency of the sound wave is perceived by the auditory organs as the pitch of the sound. Each harmonic of a sound wave is called its **tone**. The higher the pitch, the higher the frequency of the wave. The tone corresponding to the lowest frequency of the spectrum of a complex sound is called the **fundamental** tone. The tones corresponding to the other frequencies of the spectrum are called **overtones**. Overtones whose frequency is a multiple of the fundamental frequency are called harmonics, with the fundamental tone with frequency v_0 called the first harmonic, the overtone with frequency $2v_0$ called the second harmonic, etc. Sound waves whose frequencies are between 16 and 20,000 Hz are perceived by the human ear and are called **audible**. Musical or so-called tonal sounds occupy a special place among audible sounds. Musical sounds form a discrete, linear spectrum of frequencies v_t at a certain point in the spectrum of audible sounds.

Musical sounds correspond to periodic or almost periodic vibrations. The entire spectrum of musical sounds is called a **scale**, which is divided into **octaves, intervals** in which the ratio of frequencies is 1 to 2. The octave is conventionally divided into seven parts, the frequency of each part defines the corresponding note from **"C"** to **"B"**.

The number and intensity of the overtones that make up a complex musical sound determines its timbre. Sounds that form a continuous spectrum are called noise.

The measure of the strength of the auditory sense is the **loudness** of the sound, which depends on its effective pressure p_{ef} and frequency. The minimum effective pressure of a sound at which it is still perceptible to the hearing organs is called the hearing threshold. The human hearing threshold is considered to be $2 \text{-} 10^{-5}$ N / m^2 at a frequency of 1000 Hz.

The **sound pressure level** of a sinusoidal sound wave is the

$$L = 2klg\frac{p_{эф}}{p_0}$$

where k is the proportionality coefficient.

The sound pressure level is measured in decibels, and the volume level is measured in phones. 1 phon is a loudness value equal to a sound pressure level of 1 dB.

Sound waves with frequencies $v < 16$ Hz are called **infrasound** and those with frequencies $v > 2 \cdot 10^4$ Hz are called **ultrasound**. In general, ultrasound usually refers to sound waves in the range of $2 \cdot 10^4 - 10^{14}$ Гц.

Sounds between 20 Hz and 200 kHz are called low-frequency ultrasound. The emitter of the low-frequency spectrum of ultrasound is, for example, a siren. Magnetostrictive ultrasonic emitters up to 200 kHz operate on the principle of oscillations in ferromagnetics. These vibrations occur when ferromagnetics are magnetized in a periodically changing magnetic field (ultrasonic magnetostrictive vibrators). Piezoelectric ultrasonic emitters are used to generate ultrasound with frequencies up to 50 MHz. The principle of operation is similar to that of

magnetostrictive vibrators and results from the oscillation of a piezoelectric plate in an alternating electric field.

As in the case of light at high frequencies, narrow directional beams of ultrasound (ultrasonic beams) can be obtained. The directionality of the beams is the higher the closer they are to coherent waves. Ultrasonic beams behave similarly to light waves. They are subject to reflection, refraction, focusing, etc. The laws of ray optics are therefore applicable to them.

Mirrors and lenses of various shapes are used to change the direction and focus the ultrasonic beams. When making lenses and mirrors, the acoustic resistances of the materials from which they are made and of the medium are primarily taken into account.

Amplitudes of velocity, acceleration of medium particles, as well as sound pressure amplitudes in ultrasonic waves are many times greater than in ordinary sounds. In this regard, ultrasound has a crushing (destructive) effect, and in liquids causes cavitation (formation and disappearance of internal discontinuities of solid structure, having the form of tiny bubbles).

Ultrasound is intensely absorbed by liquids and gases. In engineering, ultrasound is used for sonar (similar to radar). Ultrasonic is also used to detect internal product defects (non-destructive testing) in the form of cracks, cavities, etc. Non-destructive testing is based on the scattering of ultrasonic waves from the surfaces of defective places.

Ultrasound is also used to form emulsions, suspensions, removal of oxide films, degreasing surfaces, drilling, grinding, polishing products made of metals and plastics.

Ultrasound is also widely used in medicine, chemistry, biology and other industries. Ultrasound, in particular, serves as a means of communication and biolocation of dolphins and bats.

6.6.5.4. Surface acoustic waves (SAW)

Elastic waves propagating along the free surface of a solid or along the boundary of a solid with other media (vacuum, gas, liquid,

solids) are called **surface acoustic waves**, SAW. A distinction is made between SAWs with **vertical** and **horizontal polarization**. For SAWs with vertical polarization, oscillations of medium particles occur in the plane perpendicular to the boundary. For SAWs with horizontal polarization, oscillations of the medium particles occur in a plane parallel to the boundary. The simplest SAWs with vertical polarization are **Rayleigh waves** propagating along the boundary of a solid body with vacuum or a rarefied gas medium (Fig. 6.31).

Waves with vertical polarization propagating along the boundary of two solid media are called **Stoneley waves**, which consist essentially of two Rayleigh waves.

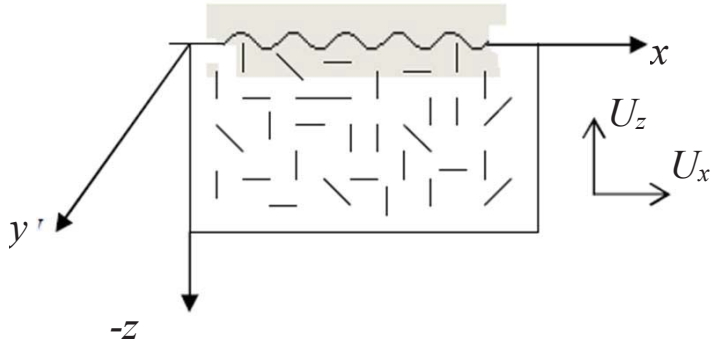


Figure 6.31.

Waves with horizontal polarization are called **Love waves**. Love waves propagate at the boundary of a solid half-space with a solid layer. These waves are transverse waves. At the boundaries of crystals, due to their anisotropy, the movement of particles becomes significantly more complicated. A crystalline medium is also often **piezoelectric** with a **piezoelectric effect** (see section 6.6.4.3). The piezoelectric effect can be **direct** or **converse**. The direct effect is called the piezoelectric effect, when electric charges appear on the crystal surface under the action of deformation. The piezoelectric effect is called the **converse effect**, when a change in the direction of the electric field applied to the crystal causes it to deform. Due to the piezoelectric effect, it is possible for SAW to interact with electromagnetic fields, including the fields of conduction electrons. On crystal planes, SAW can exist on a free surface, in the absence of a solid layer. The second medium in this case is air or vacuum.

In the general case, SAWs arise under the action of the strain forces of full compression or tension, which, as we know (see Section 2.9.6), are expressed through the divergence of the symmetric shear stress tensor T_{ij}, *where* $i,j = x,y,z$. If we denote the vector of generalized deflection by $\vec{u}$, then the equation of motion of particles on the surface of a solid body (the wave equation of the process) can be written in the form

$$\rho = \frac{d^2\vec{u}}{dt^2} = div\vec{\tau}$$

In component form, this equation is written by the following system

$$\rho = \frac{d^2u_i}{dt^2} = \sum_j \frac{\partial \tau_{ij}}{\partial x_i}$$

The piezoelectric effect is accounted for by augmenting the above system of equations with a system for the electric field, that is, the equation for the vector of electric induction $\vec{D}$ and the potential φ

$$\vec{D} = -grad\varphi.$$

It can be shown (we will accept this without proof) that the system of equations describing the wave process in this case admits a solution in the form of a plane wave with wave number $k = \omega/\upsilon$, propagation velocity υ, directional cosines b_i, displacement u_i and potential φ, namely

$$\left.\begin{aligned} u_i &= \alpha_i e^{k(b_i x_j - \upsilon t)} \\ \varphi &= \alpha e^{k(b_i x_j - \upsilon t)} \end{aligned}\right\}$$

It is important to note that neither the electric field potential nor the field itself is electromagnetic in nature in this case, because it is caused by a mechanical wave propagating at a speed υ, which is about 10^{-5} times less than the speed of propagation of electromagnetic waves.

This property of SAWs allows them to be used in long-duration delay lines and microwave technology. Since the low propagation velocity of SAW also reduces the wavelength compared to an electromagnetic wave of the same frequency, SAW devices have much

less size and weight compared to electromagnetic devices of the same purpose.

In addition, SAW devices are located on the surface of the crystal, which makes them more robust and reliable. All this has led to a wide range of applications of SAW devices, including, in addition to delay lines, as bandpass filters, control elements of microwave generators, frequency synthesizers, correlators, intermediate frequency filters, optical and acoustic image scanners, as well as signal processing devices, etc.

The propagation of SAW on the surface of the sound tube allows to easily carry out the supply and withdrawal, transmission and processing of signals, as well as - to manufacture electroacoustic transducers of SAW by methods of planar technology.

Because of these properties, SAW devices are comparable and well matched to microchips.

6.6.6. Waves on the liquid surface

On the free surface of the liquid, special waves are formed, which are sometimes, not quite correctly, called ***surface waves***. The fact is that other waves, such as the acoustic waves we discussed in the previous paragraph, can also arise on the surface of bodies. The most well known are surface waves, which are formed on the free surface of reservoirs, oceans, seas, lakes, rivers, etc. Deviation of particles of free surface of liquid from the equilibrium occurs under the action of disturbing factors (wind, movement of bodies on the surface of the liquid, etc.). The resulting liquid deformation is restored by two forces acting simultaneously on the liquid surface, ***gravity and surface tension***.

In a particular case, only one of these forces may apply. At a sufficiently deep depth, waves arise mainly under the influence of gravity. These waves are in the long-wave part of the spectrum. At relatively shallow depths, waves arise under the predominant influence of surface tension forces. These are short waves, which most often have the appearance of ripples, and are called **drip waves**.

The liquid in the surface waves remains stationary on average, but the waves themselves are moving. Unlike all other known waves, surface waves are neither transverse nor longitudinal, but a mixture of both. This is a consequence of the fact that liquid is by nature incompressible. As a result, ridges and troughs on its surface are not formed by the upward and downward movement of its particles, but by the inevitable flow of the fluid and the movement of its particles along the surface to the sides. Since the forces of surface tension and gravity resist this movement, a wave occurs. As a result of the fluid flowing out of this place, it forms a trough, which alternates with the ridge that occurs here when it flows in. The surface particles of the liquid, thus, under the action of the above forces, make a complex circular motion (Fig. 6.32).

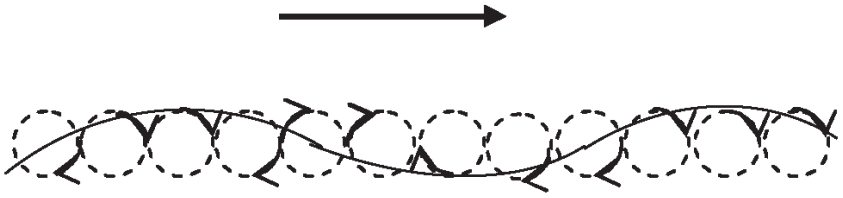


Figure 6.32.

You can easily see this by looking at objects that are close to the incoming wave.

The surface wave propagation velocity can be determined by the method of the theory of dimensions with subsequent refinement of the values of constant coefficients by experiment (see Problem 11).

Phase velocities for waves of different nature, arising respectively under the action of predominantly gravity *(m)*, surface tension *(n)*, or under the combined action of both factors, are as follows

$$v_{m\phi} = \sqrt{\frac{g\lambda}{2\pi}};\ v_{n\phi} = \sqrt{\frac{2\pi\sigma}{\rho\lambda}};\ v_{\phi} = \sqrt{\frac{k\sigma}{\rho} + \frac{g}{k}}$$

where

g is the acceleration of gravity;

λ is the length of the arising wave;

σ is the surface tension of the liquid;

ρ is the density of the liquid;

k is the wave number of the wave.

The group velocities are 3/2 and 1/2 of the phase velocity, respectively.

The study of short-wave waves (ripples) showed that the surface waves do not have a sharp wave front. This is explained by the fact that due to the dispersion, which follows from the above formulas, the speed of the main waves, which consist of fine ripples and are therefore short-wave, is greater than the speed of the other waves. Therefore, these waves are surging forward, and consequently, a sharp disturbance of the water surface does not result in a delineated wave. When an object, such as a boat, moves through the water at a certain speed, since different waves move at different phase velocities due to dispersion, a rather complex picture emerges on the surface of the water.

The above formulas show that at depth, where the wave length is greater, the wave accelerates, and near the shore, where the depth is less, its speed decreases, and the wave slows down sharply. The same wave braking occurs on all shoals. As a result, when the wave hits the shore and shoals, a shock wave occurs, which has a very complex shape and is accompanied by a strong foaming of the water.

If the observer is moving relative to the water, then, from his point of view, the water, on the contrary, is moving relative to him, and the waves, which in the end are always stationary relative to him, move as if together with him. Since, however, the group velocity of long waves is greater than the phase velocity, and of short waves is smaller, ripples are always formed in front of the floating object, and long waves are formed behind it.

6.6.7. Ray optics

As mentioned above, visible light, which plays an enormous role in human and animal life, occupies a minuscule area on the electromagnetic radiation scale. The width of the spectrum of this region is in the range of 0.45 - 0.75 μm. For comparison, recall that the entire

spectrum of electromagnetic radiation occupies a region 10^{-15}-10^{5}m wide. (see Fig. 6.23). In spite of this, it is visible light waves that carry all visual information, which for humans is at least 90% of the information about the surrounding reality received throughout his life. Visual sensations are converted by the brain into a set of images of real things. They are perceived by means of the organs of vision, part of which are the eyes. It is amazing how highly sensitive the eyes are to light. Exposure to as little as 10^{-17} watts of light energy is sufficient to produce a light sensation. The sensitivity of the eye to light of different wavelengths is different, so the subjective assessment of the light power of different colors in the visual sense and the power value are not the same. The eye, in particular, has the greatest sensitivity to yellow-green (0.54 to 0.58 μm). The sensitivity of the eye to red and violet is much lower, and in the infrared and ultraviolet, as well as beyond them, it becomes zero. For this reason, the energy characteristics of light are mainly used to assess light effects. These characteristics are set using a number of physical quantities - including luminous flux, light intensity, illuminance and brightness. **Light flux** is the amount of **light energy** flowing through a given area, per unit time. **Light intensity** is defined as the amount of light flux enclosed in one steradian of the solid angle. The SI unit of light intensity is 1 **candela (1 cd),** which is equal to the unit of luminous flux in 1 steradian of solid angle and is determined with the help of a special light standard.

Luminous flux, expressed in light intensity, is measured in **lumens (lm).** 1 lm is the luminous flux emitted by a 1 cd light source at a solid angle of 1 steradian. **Illuminance** is the amount of luminous flux per unit area of the illuminated surface. The unit of illuminance, called **lux (lx),** is the illuminance of such a surface, per unit area of which a luminous flux of 1 lm falls.

Let us denote:

Ɛ is the magnitude of the light energy incident on the site o at an angle α;

Ω is the value of the solid angle, which is defined as the value of the surface cut by the cone of the light flux coming from the light source

on the surface of the sphere S of radius g, divided by the square of this radius;

I is the power of light;

F is the luminous flux;

E is the illuminance;

Then, according to these definitions, we can write

$$F = \frac{d\mathcal{E}}{dt};\ \ I = \frac{dE}{d\Omega};\ \ E = \frac{dF}{d\sigma}$$

where

$$\Omega = \frac{S}{r^2}, \qquad \text{a} \quad d\sigma = \frac{dS}{\cos\alpha}$$

Calculations show that at noon, when the Sun is in the zenith and its rays fall on the Earth at relatively small angles, the illumination of the Earth's surface in the middle latitudes is about 10^5 lux.

The above relations make sense only for so-called **point** light sources, i.e. such sources whose dimensions can be neglected compared to the distance from them to the observer. Radiation from point sources is considered isotropic (uniform in all directions).

In addition, it is assumed that light is emitted as radial beams of rays emanating from a point source, as from the center of the sphere of the light wave front. In the case of sources of finite size, they are divided into small areas, each of which can be regarded as a point source. Extended sources are also characterized by **luminance**, which is defined as the intensity of light emanating from a unit of apparent magnitude of the source surface. Mathematically, the brightness of the source L is calculated by the formula

$$L = \frac{dF}{d\Omega} dS_1 \cos\theta,$$

where S_1 is the area of the radiating area;

θ is the angle between the normal to the surface of this area and the radiation direction.

The unit of brightness of a luminous area is the brightness of a unit of its area in the perpendicular direction at a luminous intensity of 1 cd. In the SI system, the unit of brightness is called a **nit (nt)**. Optics in which it is assumed that light is a set of light rays coming from a point source or from a set of point sources is called **ray optics** or **geometric optics**.

As follows from the definition, a point light source radiates equally in all directions, as a result of which the energy of the light flux is distributed uniformly over a spherical surface whose radius is equal to the distance from the observer (observation point) to the source. The effect of light is determined mainly by the illuminance of the area, which is inversely proportional to the square of the distance of the area to the source and directly proportional to the cosine of the angle between the direction of the incident beam and the positive normal of the area. This explains the difference in sunlight and the thermal effects of light in different climates, at different times of the year and day.

Very often it is necessary to eliminate the scattering of light energy in all directions and concentrate it in a given direction. Such concentration is achieved both by using quantum generators (lasers) and by using mirrors with a complex surface, such as parabolic mirrors, etc. In addition to these devices, for the concentration of light fluxes, devices reflecting, scattering or refracting light are also used.

In the general case, when light falls on the boundary of two media, part of the light flux is reflected, part is refracted and transmitted to the next medium, and part is absorbed. A body that reflects light more strongly than surrounding bodies is perceived as a light spot on a dark background and vice versa. Bodies that do not absorb but transmit light by refracting it are transparent. If, for example, a glass plane-parallel plate with very little absorption is placed in the path of a light beam, then the light, having passed through the plate and refracted twice, retains almost entirely its direction and intensity, so that all objects beyond it are clearly visible and undistorted. As the thickness of the plate increases, the light it absorbs increases, and it gradually loses its transparency. To change the direction of the light rays, they must be directed at the curved surface.

Let a luminous object be given. Choose an arbitrary luminous point S on it . From S, as from a point source of light, rays propagate in all directions. Choose two some rays originating from point S. Let's make them intersect at some point Si and diverge again (see Fig. 6.33).

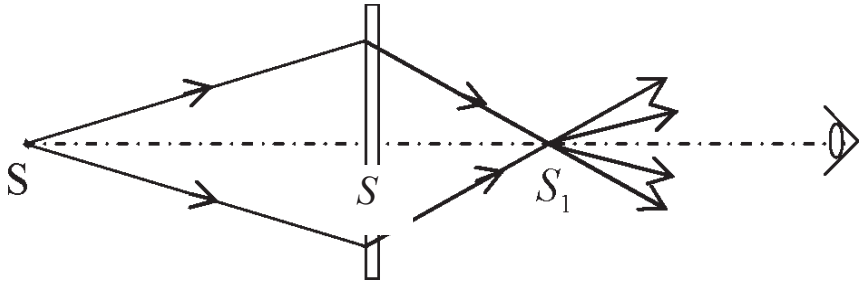


Figure 6.33.

The eye or any other light receptor outside of point S_1 will determine that the rays diverge from point S_1 as from a point source. If you put a screen or a photographic plate at this point, they will fix a luminous point. This allows us to consider point S_1 to be a **real image** of point S. We have considered two rays coming from one luminous point and intersecting at another. If we were to consider not two but three, four, and so on rays intersecting at the same point, the essence of the matter would not change. Therefore, here and hereinafter we will limit ourselves to considering only two of all possible rays. On the other hand, let us emphasize that the luminous point differs from its image because it is a real source of light and the rays from it diverge in all directions, while from the image point the rays diverge only in a limited solid angle. The above applies to all points of a luminous body, so the set of their images forms an image of the body. This image differs in that it reproduces only the distribution of amplitudes of the corresponding waves and does not reproduce the distribution of their phases. This is due to the fact that all image points, regardless of the time of their occurrence, lie in the same plane and therefore give a **flat** image of the volumetric object on the screen.

To get an image of a luminous point, you have to change the direction of the rays coming from it by making them intersect. This can be done in two ways - by refraction or reflection. Consider first the method of changing the direction of the rays by refracting them. For this purpose, a transparent plate with a curved surface of the interface between

the two media is used. This plate can be made of glass, quartz, transparent polymer, etc. In this case, at least one surface of the plate out of the two must not be flat. Such a plate is called a **lens**. The easiest way to make a lens surface is to make it spherical, although sometimes it is made cylindrical, parabolic, etc. Lenses with two different surfaces are often used, including biconvex, biconvex, convex-concave, convex-flat, and concave-flat (see Figure 6.34).

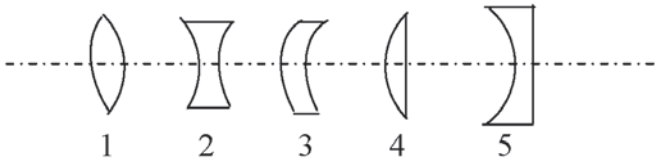


Figure 6.34.

Any real lens has a certain thickness, but this thickness is made as small as possible with respect to the radius of curvature of its surface, so that it can be neglected. Such a lens is usually called a **thin lens**. Let's take a thin spherical biconvex lens of O_1O_2 thickness as an example (Fig. 6.35). C_1 and C_2 are the centers of the spherical surfaces of the lens. If $O_1O_2 << O_1C_1$ and O_2C_2, then the points O_1 and O_2 can be considered to coincide with the point O, which is called the **optical center** of the lens.

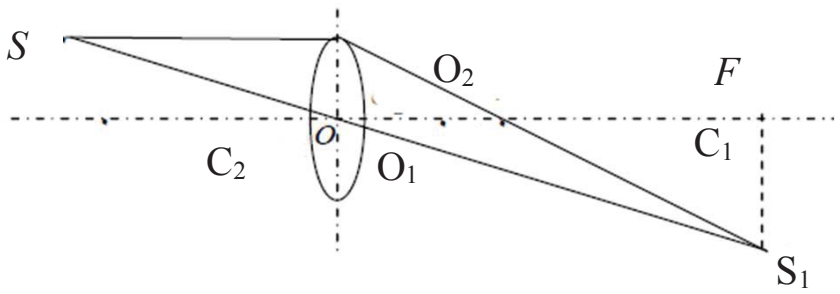


Figure 6.35.

Parts of spherical surfaces near the optical center are approximately parallel, so when a ray passes through the optical center, it practically does not change its direction, but only shifts parallel to itself. Since the shift is relatively small, we can assume that any beam passing through the optical center is not refracted. The straight lines passing through the optical center of the lens are called **optical axes**. The optical axis passing through the centers of the spherical surfaces of the lens is called the **main axis**, the other axes are called **secondary** axes. Based on

the law of refraction and geometrical considerations, it can be shown that all rays of a beam parallel to the main optical axis of a thin lens will intersect at the same point *F* on the main optical axis after being refracted twice. This point is called the focus of the lens, and the distance $OF = f$ is called the **focal length**. On the contrary, the rays originating from a point light source placed in focus will refract out of the lens as a parallel beam. The focus of a lens always lies on the same side as the center of its refractive surface. In the case of a lens with a convex surface, the focal point, where the rays falling on the lens surface and parallel to its main optical axis gather, lies on the back side of the lens. Such a lens is called a **converging lens**. The focus of a lens with a concave surface is on the same side as the rays, so they are not converging but diverging when refracted in the lens (see Figure 6.36).

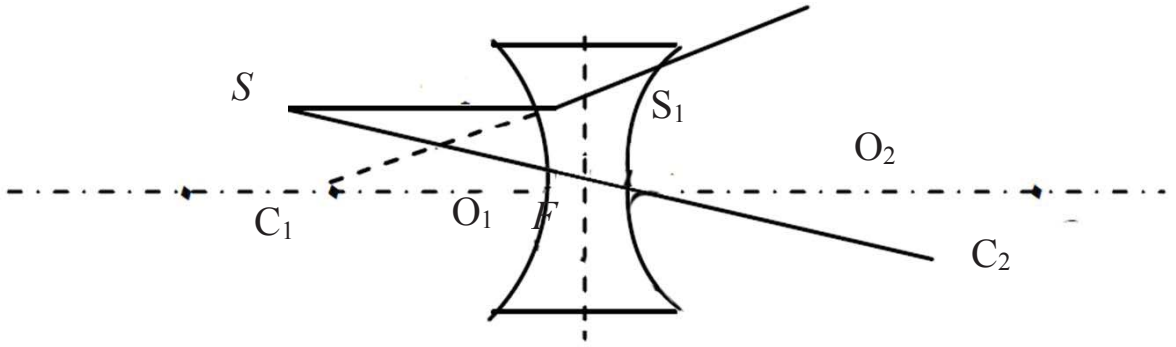


Figure 6.36.

Such a lens is called a **diverging lens** and its focus is **imaginary.** The image in the diverging lens appears at the intersection point not of the refracted rays themselves, but of their imaginary extensions. This is why it is called **imaginary** rather than real. The real or imaginary image of any luminous point of an object, and consequently of the object as a whole, is easily constructed by considering the known course of two rays emanating from that point and finding the point of their intersection, after refraction or the point of intersection of their imaginary extensions (see Figs. 6.35 and 6.36).

As the first ray, we choose a ray passing along the corresponding secondary axis and not changing its direction in this connection, and the second - a ray parallel to the main optical axis, passing after refraction through the focus of the lens. The focal distance can be determined

experimentally by the luminous point arising on the main optical axis from a beam of rays parallel to that axis directed at the lens. For a thin lens and narrow beams of light incident on the lens, the value of the focal length is determined by the properties of the lens, namely the index of refraction of the substance *n* of which it is made, and the radii of curvature of its surfaces r_1 *and* r_2. The calculations give the following relationship for the focal length

$$f = \frac{1}{n-1}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)$$

The inverse of the focal length is called the **optical power**. Optical power is measured in dioptres. 1 **dioptre** is the optical power of a lens whose focal length is 1m. The focal length and optical power of the converging lenses is positive, that of the diverging lenses is negative.

To construct an image of a luminous point lying on the main optical axis, it is necessary to first construct an image of any point lying on a perpendicular to the main optical axis, reconstructed from this point. The resulting image is projected on the main optical axis.

From Fig. 6.35 we can also obtain the so-called lens formula, which relates the focal length and the distances from the object and its image to the lens, provided that the rays under consideration form a small angle with the principal optical axis, as well as the lens magnification formula.

$$\frac{1}{f} = \frac{1}{a} + \frac{1}{a_1}$$

$$D = \frac{a_1}{a}$$

where

a is the distance from the object to the lens;

a_1 is the distance from the image to the lens;

D is the magnification.

From the above lens formula, we see that at

$$a \to \infty,\ a_1 = f; a = f,$$
$$a_1 \to \infty; a = 2f,\ a_1 = 2f; a > 2f; a > 2f,\ a_1 > f;$$
$$f < a < 2f,\ a_1 > 2f; a < f, -\infty < a_1 < 0$$

A negative value of a1 means that the image is imaginary.

It has already been said above that real and imaginary images can be obtained not only by refraction of rays with lenses, but also by reflection with concave, flat and convex mirrors. For mirrors we can use the lens formula, and for spherical mirrors of radius *r* the focal length of the mirror is

$$f = \frac{r}{2}$$

The focus of a concave mirror is real and the focus of a convex mirror is imaginary. Depending on the location of the object, the image in the spherical mirror can be real or imaginary. The image in a flat mirror is always imaginary.

Lenses and mirrors are used in science and technology to make optical devices such as microscopes, projection devices, cameras, glasses, binoculars, telescopes, etc. The simplest of these devices is a magnifying glass, which is used to magnify the image of the object in question.

The principle of the lens is also used by Nature in the organs of vision, the eyes.

Optical systems are designed to look at objects that are very small and far away from us. They are characterized by magnification or sensitivity and resolving power.

Optical instruments are discussed in detail in the technical disciplines of optics. Here we note only that the possibilities of optical instruments are limited by the wave properties of light. So, for example, when looking under a microscope at objects whose dimensions are comparable to the length of the light wave, there is diffraction, that is, the light rays envelope the object in question and actually distort its image. Because of this, electron microscopes have recently been used to reduce distortions. Since the electrons correspond to the de Broglies wavelength,

which is much smaller than the wavelength of visible light, it is possible to increase the magnification by many times.

As for telescopes, the stars viewed in them are located so far away that even at high magnification they are visible at angles of view close to zero as point objects. Telescopes allow us to see only low-bright stars, as well as to extend their point images, which, when viewed with the naked eye, merge and become indistinguishable from each other.

6.6.8. Holography

Wave interference is widely used in ray optics. Interference is used, in particular, in **holography**, which is used to produce three-dimensional images of objects called **holograms**.

Holography is a method of **recording** and reconstructing a wave field based on recording the **interference pattern** that occurs when light waves reflected by an object illuminated by a light source interfere with a wave coherent with it coming directly from the source itself. The reflected wave is called the **subject wave**, and the wave from the light source is called the **reference wave**. The recorded interference pattern is called a **hologram**. When the hologram is illuminated by the reference beam a copy of the object wave reproduces the amplitude-phase spatial distribution of the wave field which was created by it when recording and an undistorted **imaginary** image of the object is formed in the same place where the object was during holography.

A diagram explaining the process of producing holograms is shown in Fig. 6.37

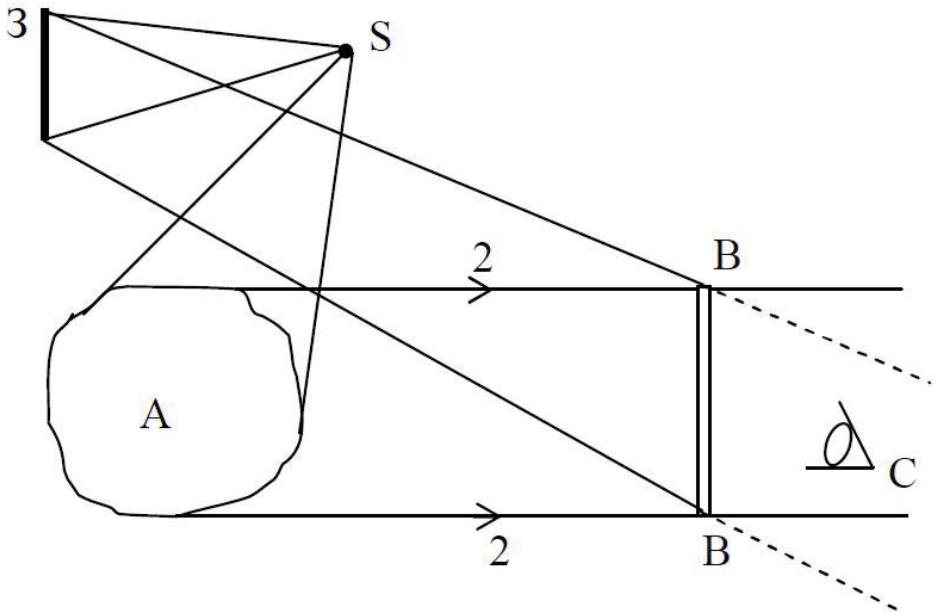


Figure 6.37.

Let the coherent light from the source S fall on the object A. The light flux reflected from the object is a certain set of reflected and scattered by the object A rays 2-2, carrying information about the object, which are called subject rays. Eye C, exposed to the indicated light flux, will see an image of the object. If a BB screen is placed in the path of the light flux, then coherent rays from all visible points of the object falling on the screen, adding up, create in different points of the screen oscillating processes different in amplitude and phase. Let's assume that the screen is transparent and is able to record information about the amplitude and phase of oscillations at each of its points. If we illuminate such a screen with light from the same source using a mirror 3, then the light wave passing through the screen is modulated by a signal whose amplitude and phase correspond to the amplitude and phase fixed in this point by the subject rays. This process is called object image reconstruction.

In other words, the light wave that passes through the screen will contain complete information about the object even in its absence. The observer's eye will see a three-dimensional imaginary image of the object in the same place. This is because a hologram, unlike a photograph, records not only the distribution of illumination, but also reproduces the phase distribution of waves reflected from different elements of the

illuminated object. The ratio of phases determines the three-dimensional nature of the image.

By moving the eye along the hologram, you can see what an object looks like from different sides of its observation. This is ensured by the fact that the interference pattern of each hologram point contains information about all elements of the object visible from this point. Another feature of the hologram is that the observer sees an image of the whole object even when the reference beam intended to reconstruct the image is not passed through the whole hologram but through a part of it. In this case the image of the object can be seen from the directions in which the object was observed from the points of the illuminated section of the hologram when this section was in its composition when it was exposed by the subject and reference beams. Fig. 6.38 shows a real scheme of obtaining a hologram.

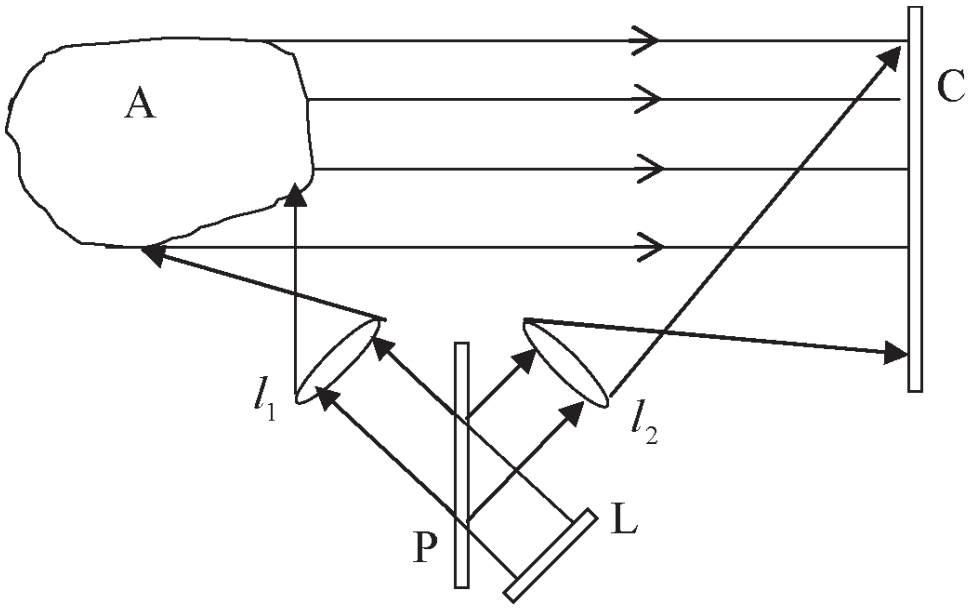


Figure 6.38.

A beam of coherent radiation from a laser *L* with the help of a semitransparent plate *P* is divided into two streams, one of which is focused by lens l_1 on the object *A*, and the second - by lens l_2 is directed to the photographic plate *C*. On the same plate the reflected and scattered on the object flux of subject rays falls, which interferes with the first, coherent to it half-flux. As a result, a hologram of the object is recorded on the photographic plate. To reconstruct an image of the object a

reference beam of light oriented relative to the hologram as it was oriented when the hologram was acquired is directed onto the hologram.

Holography was discovered in 1947 by the English physicist Dennis Gabor, but was developed in practice only after the creation in 1960 of sufficiently powerful lasers, emitters of coherent light. In Gabor's scheme the source of the reference wave and the object were placed on the axis of the hologram. In this case all three waves propagated behind the hologram in the same direction, creating mutual interference and distortion of the object image. In 1962-63, American physicists Emmett Leith and Juris Upatnieks developed a scheme with an inclined reference beam, eliminating the distortions noted.

The reconstruction of light waves recorded on a hologram creates the complete illusion of the existence of an object indistinguishable from the original, an image of which can be viewed from different angles just as a real object is viewed. These properties of the hologram are used in lecture demonstrations, in creating three-dimensional copies of works of art, holographic portraits, to study moving particles, raindrops or fog and, respectively, tracks of elementary particles in bubble chambers, creating holographic movies and television. Holographic methods are also used in interphotometry to detect changes in an object in the form of various deformations or changes in its optical properties, etc. Holography is also used to store and process information. The hologram can be produced not only by optical method, but also calculated by computer. Such a virtual digital hologram is used to produce three-dimensional images of non-existent objects. In addition, holography is also used to image an object acquired using acoustic waves. The principle of acoustic holography is similar to that of optical holography. First the standing wave field (hologram) formed by the interference of the subject and reference sound waves is recorded, and then either the original image of the subject or the structure of the sound field at some distance from the subject is reconstructed from the recording. Different methods of visualization of sound fields are used to make an acoustic hologram visible.

Recently, there have been attempts to create theories of the holographic nature of the world. This is how, for example, we think the

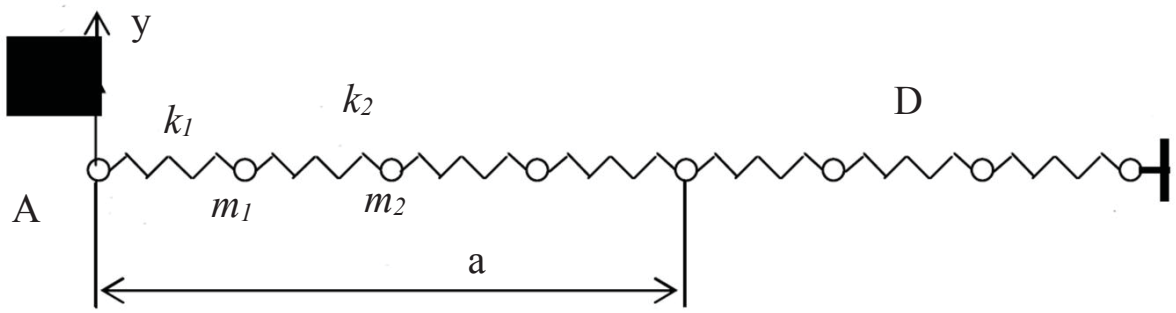
mechanism of vision can be explained. It is known that according to the theory of ray optics, a real image of an object is formed on the retina, which is then transmitted by nerve fibers to the visual center of the brain. However, why we then see this image as three-dimensional and not inside the brain, but outside the brain, in the very place where the object in question is located, is not entirely clear. It seems as if we are looking at the actual image of an object, recorded in the brain, with some kind of internal vision. This phenomenon, however, can easily be explained in terms of holography, given that a hologram appears on the retina, and we therefore, as mentioned above, see an imaginary three-dimensional image of the object in the very place where the object is located. However, in this case it is necessary to assume that there is a mechanism in the visual center of the brain which erases the hologram when all elements of an object leave the field of view of the eye. This approach is consistent with the general considerations made by Stanford University neurophysiologist Karl H. Pribram. Pribram, in particular, suggested that the brain's ability to transform perceptions of the senses into a coherent image corresponds to the principles of holography. The brain, according to Pribram, seems to probe the holographic information previously recorded by the same flows of radiation coming from the senses. Even more generalized is the "holographic principle" formulated by the Dutch physicist from Utrecht University Gerard 't Hooft, according to which all information contained in some area of space can be represented as a certain hologram registered at the border of this area.

6.6.9. Typical problems on wave processes

Problem 1. The free end of a horizontally stretched elastic cord, the other end of which is fixed stationary, is oscillated in the vertical plane with a small amplitude B and period T. What process occurs in the cord? Explain why. How does the displacement of point D of the cord, which is on the right of its free end at a distance a, depend on time?

Solution. An elastic cord consists of molecules bound together by intermolecular (elastic) interactions (bonds) (see figure). In the figure,

molecules of mass m_i are represented by circles, and the bonds between them are represented by elastic springs of stiffness k_i:



Let the free end A of the cord (a particle of mass m_1) be oscillated with small amplitude and period T along the y-axis so that

$$y_A = B\cos\frac{2\pi}{T}t$$

The natural oscillations of a particle of mass m_1 are rapidly damped and can be neglected. The oscillations of particle m_1 cause shear deformation in the cord, which is transmitted by elastic bonds from one particle to another with some speed u (the speed value is determined by the degree of elasticity of the cord, so that the higher the elasticity, the greater the speed). As a result, a transverse wave occurs in the cord (particles oscillate in the direction perpendicular to wave propagation). Perturbations from point A to point D will come with a time delay Δt

$$\Delta t = \frac{a}{\upsilon}$$

Therefore, at point D after time Δt there will be the same oscillations as at point A, but lagging behind them in phase so that

$$y_D = B\cos\frac{2\pi}{T}(t-\Delta t) = B\cos\frac{2\pi}{T}\left(t-\frac{2\pi}{\lambda}a\right)$$

Since

$$\lambda = \upsilon T$$

then

$$y_D = B\cos\left(\frac{2\pi}{T}t-\frac{2\pi a}{\upsilon T}\right) = B\cos\left(\omega t-\frac{2\pi}{\lambda}a\right)$$

where λ is the length of the elastic transverse wave propagating along the cord. Since

$$\frac{2\pi}{\lambda} = k$$

where *k* is the wave number, then

$$y_D = B\cos(\omega t - ka)$$

The wave arising in the cord is a running wave and covers the entire cord. The excited particle returns to its original state of equilibrium, and its oscillations are resumed again and again.

After the wave arrives at point C, it will be reflected. As a result, two coherent waves will arrive at an arbitrary point x of the cord, namely the traveling wave y_T and the reflected wave y_R.

$$y_R = B\cos(\omega t - kx)$$

$$y_R = B\cos(\omega t + kx)$$

When these waves are added together, a standing wave is formed

$$y = y_T + y_R$$

namely

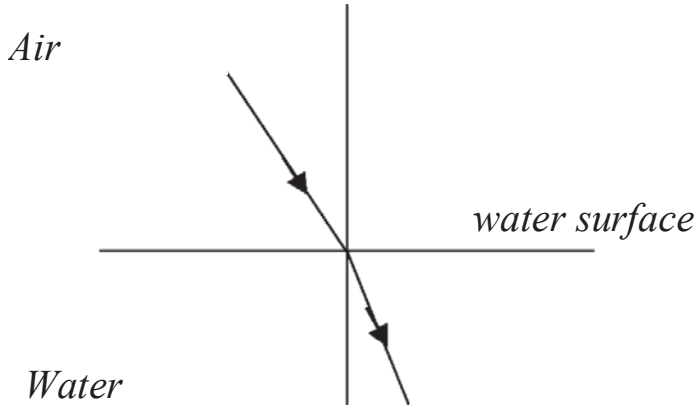
$$\mathrm{y} = B[\cos(\omega t - kx) + \cos(\omega t + kx)]$$

or

$$y = 2B\cos kx \cdot \cos \omega t.$$

In this wave, each point makes harmonic oscillations, and the amplitudes of the points of the cord ($2B\cos kx$) change depending on their coordinate x according to the same harmonic law as the oscillations of the points in time.

Problem 2. A sound wave at normal temperature falls on the interface between air and water. How does the frequency and length of the refracted wave change? Same for light (see figure)



Solution. The incident wave, hitting the points of the interface, excites secondary waves of the same frequency in them, which then propagate in the water unchanged. Thus, the frequency of the wave does not change during the transition from air to water (refraction).

The speed of propagation of sound waves in air and water are respectively

$$\upsilon_{water} = \frac{1500m}{s};\ \ \upsilon_{air} = 340\ m/s$$

The wavelength is

$$\lambda = \upsilon T,$$

It is, in this way,

$$\frac{\lambda_{water}}{\lambda_{air}} = \frac{1500T}{340T} \cong 4{,}4$$

The length of the sound wave in water will increase 4.4 times.

For visible white light, the refractive index of water is 1.33, and for air it is 1.00029. The **refractive index** of light in a medium (absolute refractive index) is the value

$$n = \frac{c}{\upsilon},$$

where

c is the speed of light in a vacuum;

υ is the speed of light in the medium.

For water $n \cong 1{,}33$, therefore

$$\frac{c}{v} = 1{,}33$$

For air $n \cong 1{,}0029$, and

$$\frac{c}{v} = 1{,}00029.$$

Let us divide each of these expressions by the period *T* of the light wave

$$\left.\begin{aligned} \frac{c}{v_{water}T} &= \frac{1{,}33}{T} \\ \frac{c}{v_{air}T} &= \frac{1{,}00029}{T} \end{aligned}\right\}$$

from which

$$\left.\begin{aligned} \frac{c}{\lambda_{water}} &= \frac{1{,}33}{T} \\ \frac{c}{\lambda_{air}} &= \frac{1{,}00029}{T} \end{aligned}\right\}$$

We divide the first equality by the second

$$\frac{c}{\lambda_{air}} \cdot \frac{\lambda_{water}}{c} = \frac{\mathrm{T} \cdot 1{,}33}{1{,}00029T}$$

then

$$\frac{\lambda_{air}}{\lambda_{water}} = 1{,}33$$

Thus, during the transition from air to water, the sound wavelength increases 4.4 times, and the light wavelength decreases 1.33 times.

Problem 3. Calculate the natural frequencies of the air column oscillations in a 2.5 m long tube closed at both ends.

Solution. The natural oscillations of the air column are a set of standing waves. Since the pipe is closed at both ends, there are wave

nodes at these ends. The distance between two adjacent standing wave nodes is λ/2, and there must be an integer number of half-waves on the length l of tube, then

$$l = m \cdot \frac{\lambda}{2}$$

from which

$$\lambda = m \cdot \frac{2l}{m}$$

On the other side,

$$\lambda = cT = \frac{c}{v}$$

where *T and v* are, respectively, the period and frequency of natural oscillations, whence

$$v = \frac{c}{\lambda} = \frac{cm}{2l}, \qquad where\ m = 1,2,3, \ldots$$

The oscillation of the air column excites sound waves, and for sound

$$c = 340\ m/s$$

therefore

$$v = \frac{340}{2 \cdot 2{,}5} m$$

or

$$v = 68 \cdot m\ Hz, \qquad where\ m = 1,2,3, \ldots$$

Fundamental tone $v_0 = 68\ Hz$

1st overtone $v_1 = 136\ Hz$

2nd overtone $v_2 = 204\ Hz$

etc.

Problem 4. To transmit sound over long distances in radio engineering, an informative low-frequency electromagnetic signal (modulation) from an electroacoustic transducer, a microphone, is

superimposed on the high-frequency carrier electromagnetic signal. How many periods of electromagnetic oscillations emitted by an electromagnetic wave transmitter of length l = 100 m are contained in one period of modulating oscillations, when the sound frequency is 1200 Hz.

Solution. Electromagnetic wave length

$$\lambda_{em} = \upsilon T = \frac{\upsilon}{\upsilon_{em}}$$

where υ is the speed of propagation of electromagnetic waves in the air. Assuming that the speed in air is equal to the speed of wave propagation in a vacuum, let us assume that

$$\upsilon = \mathrm{c} = 3 \cdot 108\, m/c.$$

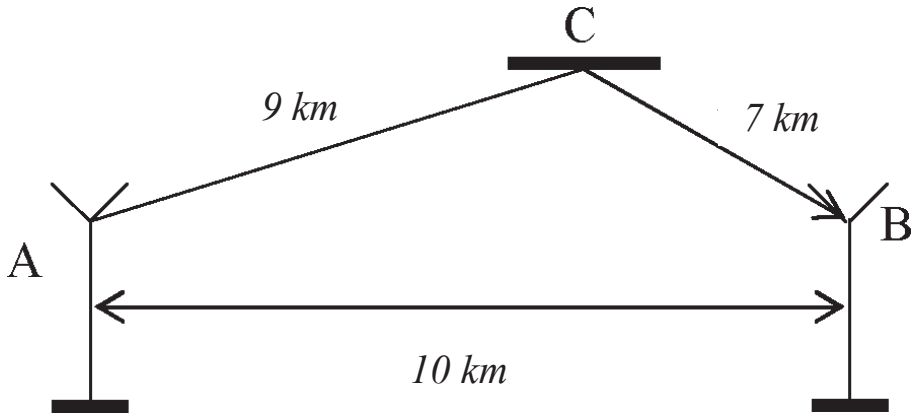
Thus, the frequency of electromagnetic waves is

$$\upsilon_{em} = \frac{3 \cdot 10^8}{100} = 3 \cdot 10^6\, Hz$$

From here

$$\frac{\upsilon_{em}}{\upsilon_{sound}} = \frac{3 \cdot 10^6}{1{,}2 \cdot 10^3} = 2500$$

Problem 5. The antenna of the television receiver *B* receives, along with the wave from the broadcasting center *A*, also the wave reflected from the object *C*. The distances are shown in the figure. What happens to the picture on the TV screen? The width of the screen is 0.5 m. Frames on the screen change with a frequency of *n* - $20s^{-1}$. The image consists of *N* = 600 lines



Solution. Since the reflected signal travels a distance of 16 km and the direct signal travels 10 km, the reflected signal will arrive with a delay $\Delta t = \Delta T/c$ and cause a blurring (splitting) of the image on the screen by the amount of

$$\Delta l = \upsilon \Delta t$$

where υ is the speed of the electron beam on the screen.

The length of one line is equal to *l* (by screen width), and the length of *N* lines is equal to *Nl*. This electron beam travels this way in time *1/n*, so

$$\upsilon = \frac{Nl}{1/n} Nnl$$

from which

$$\Delta l = Nnl\Delta t$$

Substitution yields

$$\Delta l = 600 \cdot 20 \cdot 0{,}5 \frac{6 \cdot 10^3}{3 \cdot 10^8} = 0{,}12\ m\ (12\ cm)$$

Problem 6. It is known that the radio galaxy 3C295 is distant from us at a distance R = 1.85 • 10^{10} pc (1 parsec = 3.086 • 10^{16}m). According to Hubble's law, it is moving away from us at the speed of

$$\upsilon = HR,$$

where *H* is the Hubble constant

$$\mathrm{H} \cong 75 \frac{km/s}{Mpc}$$

Determine the red shift in the spectrum of the radio galaxy.

Solution. The red shift value is determined by the Doppler effect for optical phenomena, according to which , /^2

$$v = v' \frac{\sqrt{1 - \frac{\upsilon^2}{c^2}}}{1 + \frac{\upsilon}{c} \cos\theta}$$

In our case $\theta = 0$, and $\cos\theta = 1$.

From Hubble's Law

$$\upsilon = 75\frac{km}{s}\frac{1{,}85 \cdot 10^3 Mpc}{Mpc} \cong 1{,}39 \cdot 10^8 \frac{m}{s}$$

Since it follows from Doppler's law that

$$\frac{1}{T} = \frac{1}{T'}\frac{\sqrt{1-\frac{\upsilon^2}{c^2}}}{1+\frac{\upsilon}{c}}$$

then

$$\frac{1}{\lambda} = \frac{1}{\lambda'}\frac{\sqrt{1-\frac{\upsilon^2}{c^2}}}{1+\frac{\upsilon}{c}}$$

Substituting the velocity υ considering that

$$\frac{\upsilon}{c} = 0{,}463;\ \frac{\upsilon^2}{c^2} = 0{,}214$$

gives

$$\frac{1}{\lambda} = \frac{1}{\lambda'}\frac{\sqrt{0.796}}{1.463};$$

from which

$$\frac{\lambda + \Delta\lambda}{\lambda} = 0{,}6$$

Finally

$$\frac{\Delta\lambda}{\lambda} = 0{,}4$$

Thus, the red shift in the spectrum of the galaxy, measured by the value of the relative wavelength increase, is 0.4.

Problem 7. Find an expression for the wave function describing the quantum harmonic oscillator and its energy spectrum.

Solution.

As we know, in classical mechanics a one-dimensional harmonic oscillator is a particle making harmonic oscillations under the action of an elastic force (see Section 5.1.1)

$$F = -kx$$

The potential energy of such an oscillator is

$$U = \frac{kx^2}{2}$$

and the natural frequency of oscillation

$$\omega = \sqrt{\frac{k}{m}},$$

from which

$$U = \frac{m\omega^2 x^2}{2}$$

Substitution into the time-independent Schrödinger equation gives

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + \frac{m\omega^2 x^2 \psi}{2} = \mathcal{E}\psi$$

The solution of the problem is thus reduced to the solution of the resulting second-order differential equation. To simplify the equation, let us introduce the notation

$$\rho = \frac{m\omega}{\hbar}x;\ \alpha = \frac{2\mathcal{E}}{\hbar\omega}$$

then the differential equation takes the form

$$\frac{d^2\psi}{d\rho^2} + \left(\alpha - \frac{\hbar}{m\omega}\rho^2\right)\psi = 0,$$

where

$$\psi = \psi(\rho)$$

We look for the solution of the equation in the form

$$\psi(\rho) = f(\rho)e^{\frac{\rho^2}{2}}$$

where

$$f(\rho) = \sum_{k=0}^{\infty} b_k \rho^k,$$

Substituting the solution into the original equation gives the recurrence formula for the coefficients b_k

$$b_{k+2} = \frac{(2k+1-\alpha)}{(k+1)(k+2)} b_k$$

The resulting solution, however, does not make physical sense, since it is given by a divergent series. Indeed, at $k \to \infty$ $\rho^k \to \infty$ faster than the coefficients b_k to zero. To avoid this divergence, it is necessary to limit the infinite series, that is, to require that, starting from some $k = n$, the coefficients $b_n = 0$, that is, that

$$2n + 1 - \alpha = 0.$$

Substituting the value of a gives for the energy spectrum of the oscillator

$$\mathcal{E}_n = \left(n + \frac{1}{2}\right)\hbar\omega, n = 0,1,2 \ldots$$

Problem 8. Determine, up to constant coefficients, the functional dependence of the phase velocities of surface waves on the physical quantities determining them.

Solution. We solve the problem by applying the method of the theory of dimensions. It is logical to assume that the phase velocity of gravity waves is determined by the wave number k (wavelength), the acceleration of gravity g and, perhaps, the density of water p.

Taking into account the above for the phase velocity of gravitational waves we can write

$$[v_{ph}] = [g]^x [k]^y [\rho]^z$$

or

$$LT^{-1} = [LT^{-2}]^x [L^{-1}]^y [ML^{-3}]^z$$

from which

$$\begin{cases} x - y - 3z = 1 \\ -2x = -1 \\ z = 0 \end{cases}$$

The solution of the resulting system of equations gives

$$x = 1/2;\ y = -1/2;\ z = 0.$$

It is, in this way,

$$v_{ph} = \sqrt{\frac{g\lambda}{2\pi}}$$

Similarly, for the phase velocity of droplet waves, assuming that it depends, in addition, on the surface tension coefficient o, which has a dimension [N/m], but does not depend on the acceleration of gravity, we obtain

$$v_{ph} = \sqrt{\frac{2\pi\sigma}{\rho\lambda}}$$

Problem 9. Focal length of the optical system $f = 0.3$ m, the main planes *A* and *B* of the lenses of the optical system are at a distance of 0.1 m from each other. The image of the object SN, located from the front main plane at distances: 0.2m; 0.5m; and 0.8m, respectively, must be plotted, and the linear and angular magnification must be determined for each case (see figure).

Solution. An optical system is a device consisting of several lenses arranged in such a way that their major optical axes coincide to form the major optical axis of the system. The plane perpendicular to the main optical axis that passes through the optical center of a given lens is called the main plane of that lens.

In the case of an optical system there is no single main plane, but it is easy to show that any however complex optical system can be replaced by a set of two main planes spaced at a certain distance from

each other. The points of intersection of these planes with the main optical axis are called main points. The distance a from the subject to the front main plane and the front focal length f_1 are counted from the front main point. The distance from the image b to the main back plane and the back focal length f_2 are counted from the main back point.

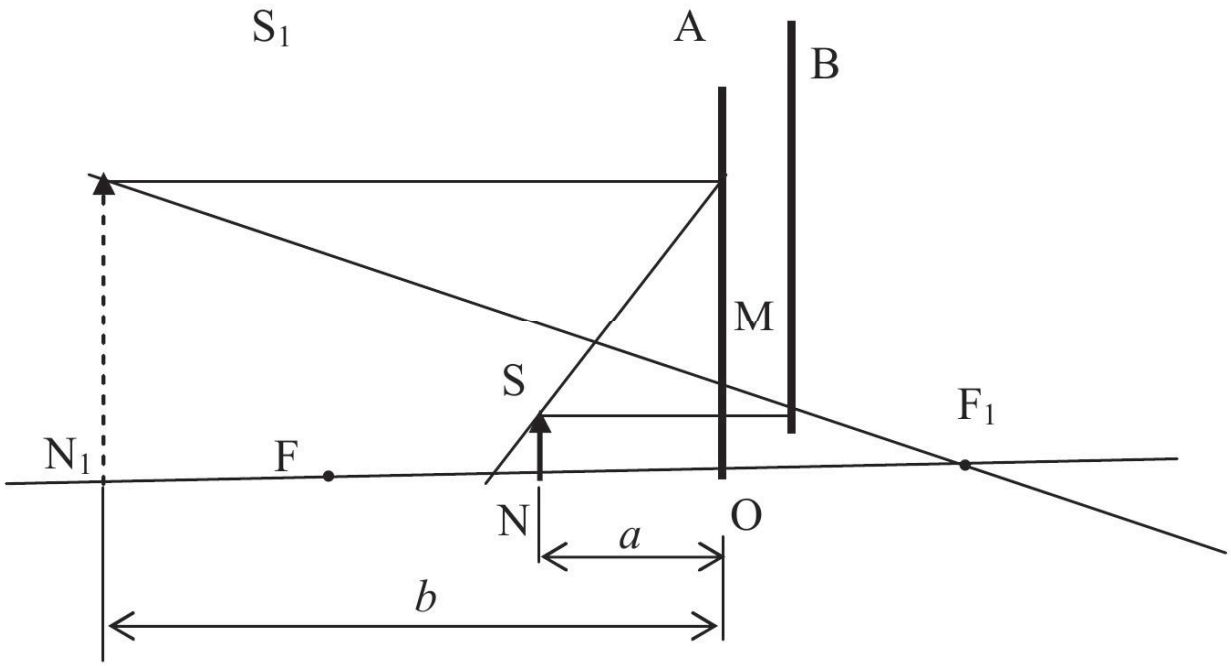


Figure 1.

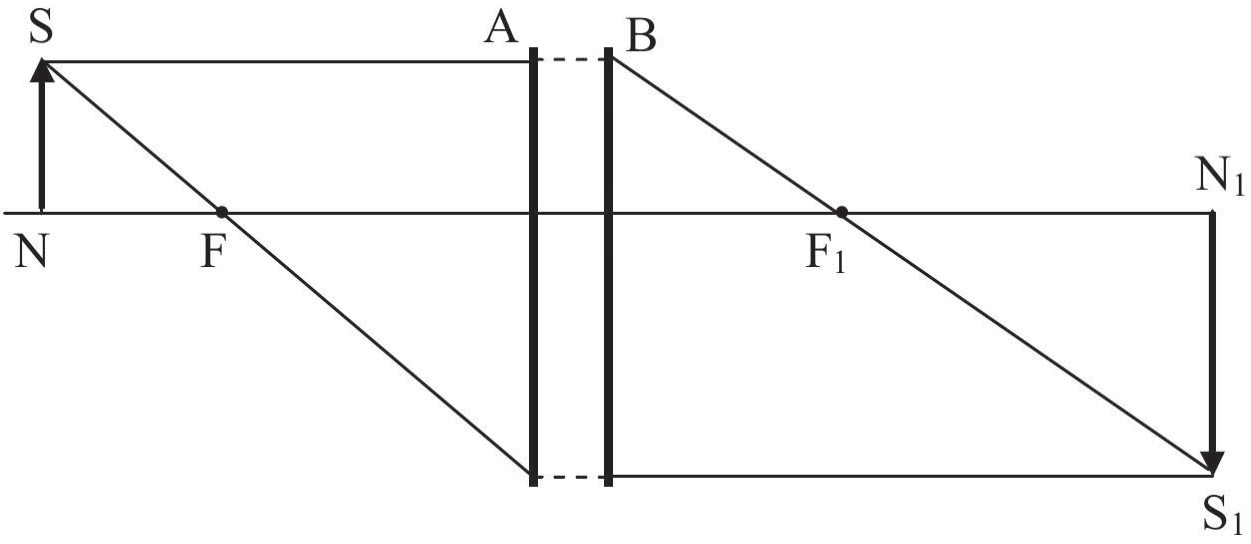


Figure 2.

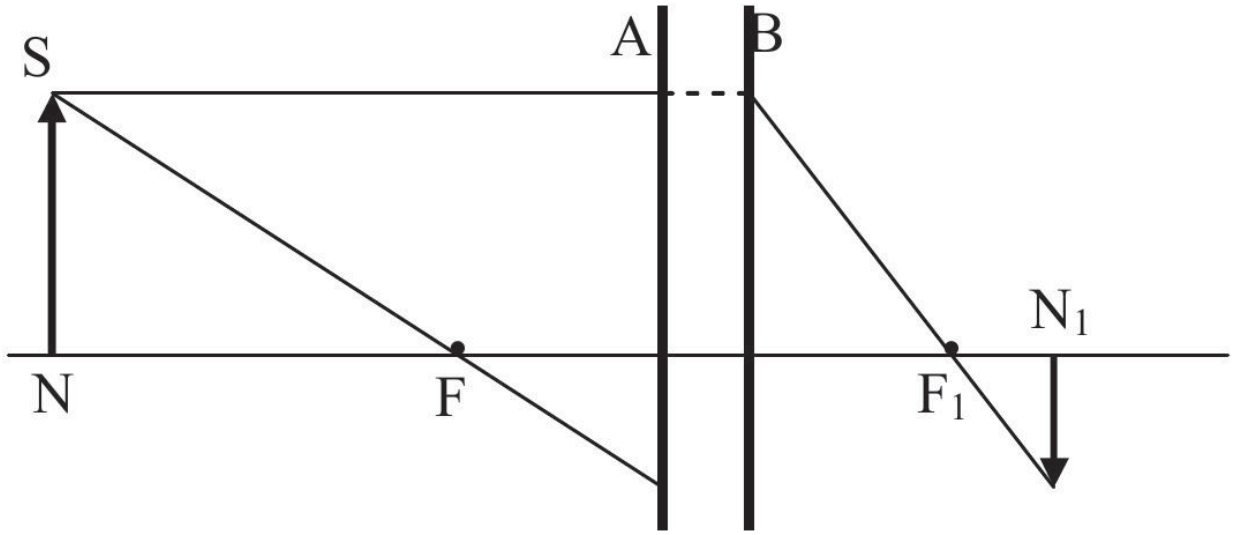


Figure 3.

Considering, for example, in Fig. 2 pairs of similar triangles SFN -MFO and $S_1F_1N_1$-$C_1F_1O_1$ provided that $f_1 = f_2 = f$, we conclude that

$$\frac{1}{f} = \frac{1}{a} + \frac{1}{b};\ \beta = \frac{N_1S_1}{NS} = \frac{b}{a}$$

$$\gamma = \frac{1}{\beta}$$

where β is a linear magnification;

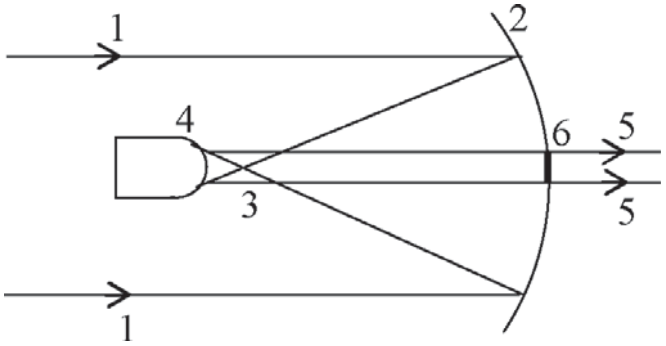
γ is the angular magnification, that is, the ratio of the tangents of the angles formed by the ray emanating from the image point lying on the main optical axis, with this axis, and the ray emanating from the subject point on the main optical axis and the axis.

From the figures and the above relations it follows that in the first case (Fig. 1) the image S1N1 is imaginary, straight, magnified, with $\beta = 3$, $\gamma = 1/3$, in the second case (Fig. 2) the image is real, inverted, magnified with $\beta = 1.5$, $\gamma = 2/3$, in the third case (Fig. 3) the image is real, inverted, reduced with $\beta = 0.6$ and $\gamma = 5/3$.

Problem 10. In the story "The Garin Death Ray" by the famous Russian writer Alexei Tolstoy, described a formidable weapon, the principle of which, in the author's mind, is based on the use of the laws of ray optics and, in particular, mirrors with a hyperbolic surface. What do you think is the scheme of such a hyperboloid and its effectiveness?

Is it possible to create a weapon that uses the same optical effect as the Garin Hyperboloid, but works on a completely different principle, using other physical laws?

Solution. The schematic of the Garin hyperboloid is shown in the figure



A parallel beam of light 1-1 from a distant natural source, such as the Sun, or from a powerful artificial source installed in the focus of the optical device is directed to the surface of a long-focus concave hyperbolic mirror 2. According to the laws of ray optics, the light beam is collected in the focus 3 of this mirror, which coincides with the focus of another, short-focus hyperbolic mirror (hyperboloid) 4.

Using a system of reflectors, a narrow beam of light coming from a point source image obtained at focus 3 is directed to the surface of the hyperboloid and reflected from it as a powerful parallel beam of light 5-5, in the small volume of which enormous energy is concentrated. Exiting the optical system through the small aperture 6, the resulting beam of light destroys and burns everything in its path.

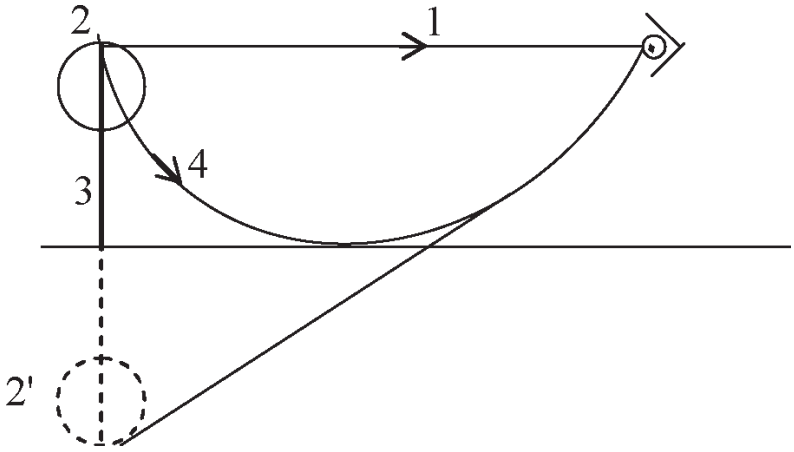
For all its wit, Garin's hyperboloid does not work for many reasons. Firstly, neither natural nor artificial point light sources exist in principle, and secondly, it is impossible to manufacture optical system elements with point focal points located exactly on the main optical axis. Moreover, it is impossible to exactly match the optical axes and focal points of the elements of the systems. In other words, it is impossible to obtain a strictly parallel beam of light from the beginning. An even more difficult, practically impossible task is the problem of keeping the rays inside a concentrated beam of small volume. This is explained, first, by the dispersion of light. The incoherence of the light emitted by any source

and the impossibility of obtaining a strictly monochromatic light beam, leads to a very rapid divergence even with a strict initial parallelism of the beams. Secondly, diffraction of light acts in the same direction, which essentially boils down to the fact that each particle of medium becomes a point source and scatters even strictly monochromatic light in all possible directions.

These difficulties of obtaining and transmitting narrowly concentrated beams of light at a distance are currently solved by using fiber-optic light guides, as well as by using quantum generators, lasers, which will be discussed in detail below, working, however, on entirely different physical principles.

Problem 11. Using the laws of ray optics, explain the well-known phenomenon of mirages in the desert, such as the mirage of a tree standing, seemingly, on the shore of a body of water. Show the course of the rays in this phenomenon.

Solution. The course of the rays is shown in the figure

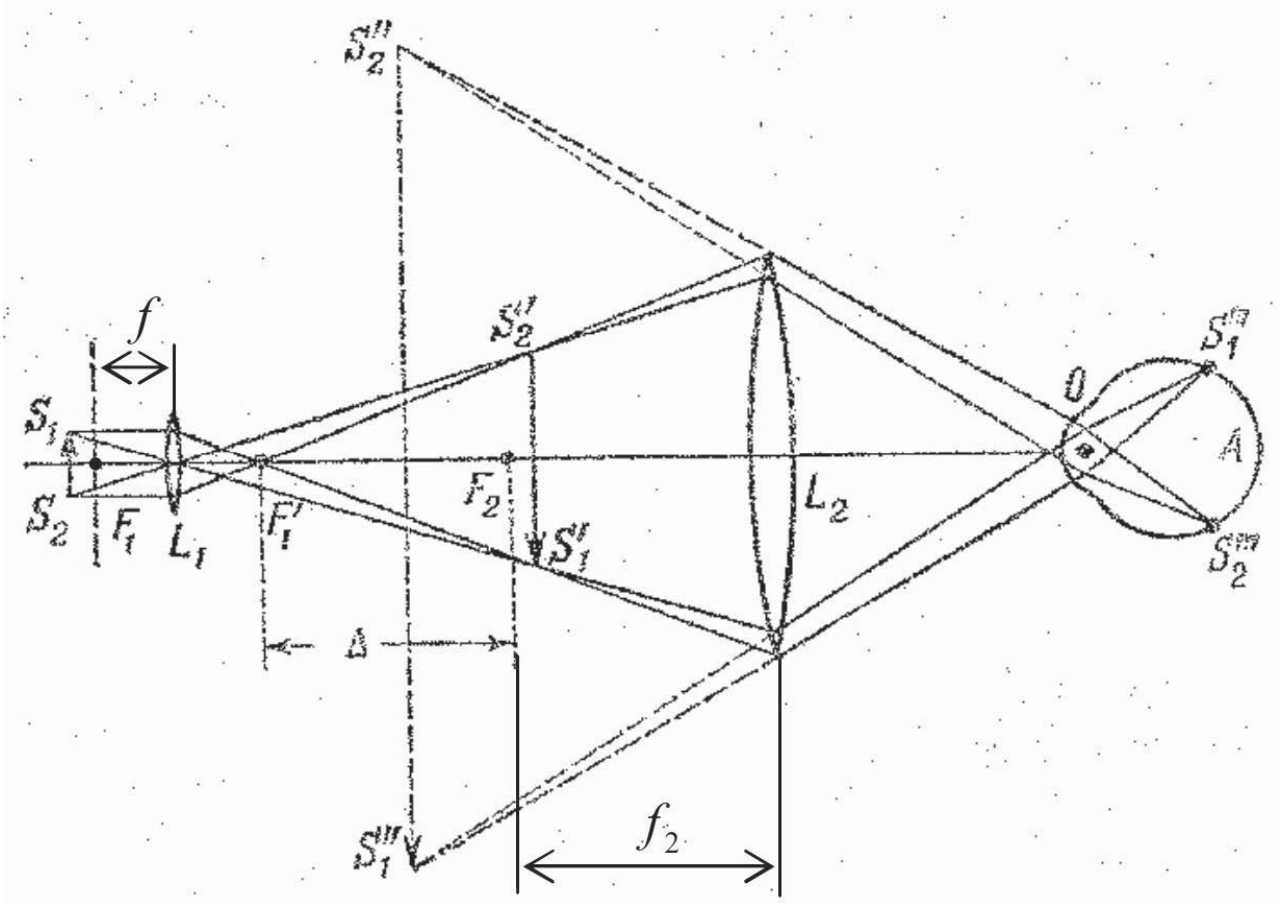


The temperature of the desert air, and consequently its refractive index, changes uniformly with changes in altitude. A layer of hot air is located directly over the hot sand. As air rises vertically upward, it gradually cools and its refractive index increases. Accordingly, the rays of light take different paths from the source to the observer in different layers of air. The ray 1 coming from the upper point 2 of a tall object 3, such as a tree, moves in the upper, relatively cooled layer of air, in a straight line, while the ray 4 coming from the lower layer, at the level of the sand surface, is curved, as shown in the figure. Crossing in the

observer's eye, these rays give him the image of point 2. Since the visual center of the human brain perceives all rays as straight, a person mentally continues the ray 4 coming into his eye in the direction of the tangent line of that ray. This ray intersects with the mental extension of the reflected from the surface of the sand ray perpendicular to it and gives at point 2/ an imaginary image of point 2. As a result, the observer will see a tree and its imaginary image, which he will naturally perceive as a reflection of the tree in the pond.

Problem 12. Draw and explain the course of the rays in an optical microscope. Calculate the magnification and focal length of the microscope if you know that the objective magnification is 70 and the eyepiece is 12.

Solution. A microscope is an optical system consisting of at least two lenses, an objective and an eyepiece. The course of the rays in the microscope is shown in the figure.



A small object S_1S_2 to be enlarged is placed at a distance slightly longer than the focal length f of lens L_1, so that its image $S'_1S'_2$ is placed near the focus F_2 of eyepiece L_2 at a distance slightly shorter than the focal length f_1. The imaginary image of the object $S''_1S''_2$ obtained with the eyepiece L_2, which is greatly magnified, is seen by the eye. Its image, according to the laws of ray optics, is perceived by the eye placed almost at infinity. Since the retina in the normal state is in the focus of the lens - the pupil of the eye, the enlarged image of the object $S'''_1S'''_2$, appears on the retina without any tension

Let us denote the magnification of the objective N_1, the magnification of the eyepiece N_2, the size of the object H, the size of the intermediate image x, the size of the final image in the eyepiece viewed by the eye H_1, the magnification of the microscope N. Then we can write that

$$N_1 = \frac{x}{H}, N_2 = \frac{H_1}{x}, N = \frac{H_1}{H}$$

from which it follows that

$$N = N_1N_2.$$

After substitution we obtain that the magnification of the microscope is

$$N = 840.$$

Show on your own that the action of a microscope is equivalent to that of a magnifying glass, from which it follows that the focal length of the microscope f_m= 0.3 m. For the microscope as an optical system (see problem 1)

$$\frac{1}{f_m} = \frac{1}{a} + \frac{1}{b}; b = Na$$

where *a and b* are, respectively, the distances of the object to the lens and the imaginary image to the eyepiece,

$$f_m = \frac{aN}{N+1}$$

Since

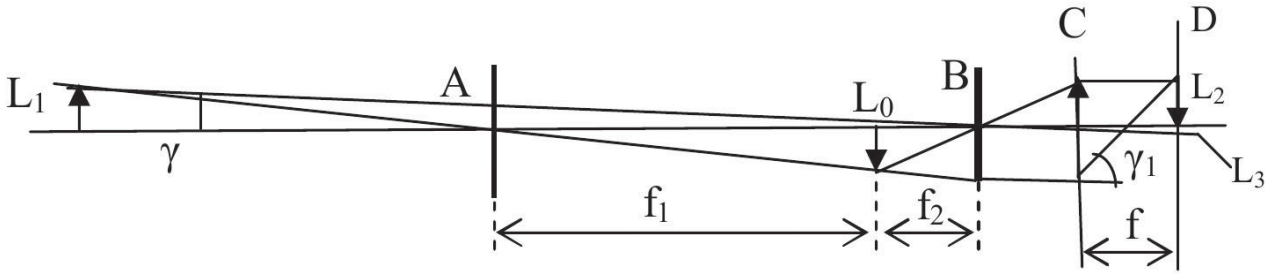
$$a \approx f; M \gg l,$$

then

$$f_m \approx f$$

Problem 13. Determine the magnification of the telescope if it is known that the focal length of objective A is $f_1 = 1.6m$, and its eyepiece *B* has 10x magnification (see figure).

Solution. A telescope is a viewing instrument used to look at very distant objects, such as celestial bodies. Like a microscope, it is an optical system containing an objective and an eyepiece (telescopes usually use long-focus concave mirrors as an objective, and short-focus converging or diverging lenses as an eyepiece). The eyepiece in the telescope is mounted so that its front focus is aligned with the rear focus of the lens. Since the object is removed almost to infinity, its greatly reduced image appears in the focus of the eyepiece. The figure shows the course of the rays in a telescope in which, for simplicity, the objective, eyepiece, and eye C are represented by converging lenses and the retina is represented by screen D



The figure shows that for a telescope it is not the linear but the angular magnification that matters. Indeed, the magnification N of the tube used as a telescope can be calculated as follows.

$$N = \frac{L_2}{L_3} \approx \frac{\gamma_1}{\gamma}$$

where L_2 is the size of the image of the object obtained on the retina (as on the screen) using a tube;

L_3 is the size of the image of an object on the retina of the naked eye.

γ is the angle at which an object can be seen with the naked eye

The figure shows that

$$L_0 \approx f_1\gamma = f_2\gamma_1$$

where

f is the focal length of the pupil of the eye;

L_0 is the size of the image of the object in the lens;

f_2 *and* f_1 are focal lengths of the eyepiece and lens, respectively.

Substitution yields

$$N = \frac{f_1}{f_2}$$

Since the eyepiece is an ordinary magnifying glass, the magnifying glass formula is valid for it, that is

$$f_2 = \frac{0{,}25}{10}$$

therefore,

$$N = 64.$$

Problem 14. Many fiction writers often turn to the image of the invisible man in their works. It is usually a talented scientist who, using the laws of ray optics in a witty way, has figured out a way to make it so that he himself sees everyone, but that no one sees him at the same time. The question is what way the invisible man came up with and whether it is possible in principle to come up with such a way.

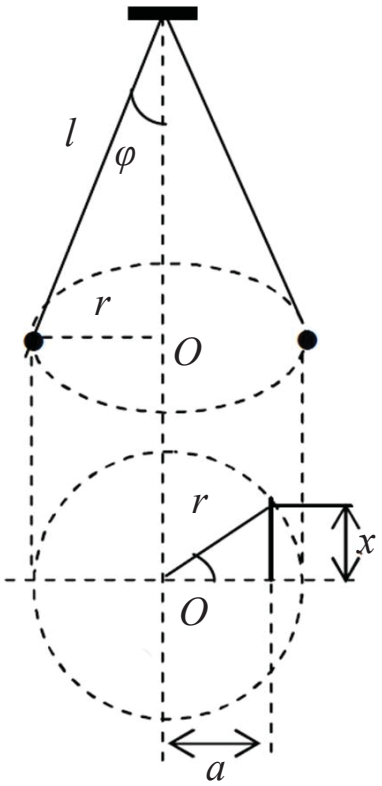
Solution. Any transparent object can be made invisible by immersing it in a transparent medium. For this it is enough to give it the same refractive index as the medium. In this case, the rays of light on the surface of its interface with the medium and in further movement inside the body will not experience changes compared to their movement inside the medium, and the observer will not be able to distinguish this body from the medium. If a body or medium is opaque, it or the medium will absorb more rays than it will reflect (compared to a transparent medium or body), and the body will either stand out against the medium or become completely invisible. In this case, a person can completely cover himself

with clothes impregnated with a composition whose coefficient of refraction is equal to the coefficient of refraction of the air. He can cover his eyes with a transparent material with the same properties. Since a human being absorbs more and reflects less rays than air, he will stand out against the background of the air environment as a dark silhouette, that is, he will remain visible. Animals have come up with a different way of doing this. They hide in an opaque, often colored medium, covering their bodies with a compound with the same coloring and light reflection pattern as the medium. This method of masking borrowed from the animals also by the military. It is, however, imperfect and stops working at close distances, so it is completely unsuitable for invisibility. This is due not only to the fact that it is impossible to accurately reproduce the reflection structure of the medium, and the difference between the body and the medium with decreasing distance to the observer becomes more and more noticeable. More significant is that the invisible man cannot hide his eyes in any way. Indeed, if he covers his eyes with a compound whose index of refraction is equal to the index of refraction of the medium, he himself goes blind, because the optical system of his eyes (their refractive power) stops working. An invisible man can, of course, hide in an opaque environment, for example, by diving deep under water, or by hiding behind some obstacle where the light reflected by his body will not reach the observer. But in this case, too, for the same reason, he will not see anyone unless he uses a periscope, that is, a device that penetrates the surveillance environment. And yet there is a way to become invisible, which, incidentally, is widely known. It is enough to go into a dark room and through a window in that room to observe those who are in any other illuminated place or in the daytime on the street. In this case, the observer himself not reflecting visible light, which is simply not present in the dark room, quietly observes everyone who is in a bright place and reflects the light falling into his eyes. They are all visible because their reflectivity is different from that of the environment. However, none of this has anything to do with science fiction anymore. However, the invisible man, being a science fiction, also has nothing to do with science, because neither now nor in the future it can in principle become a reality.

Appendix to Part 4

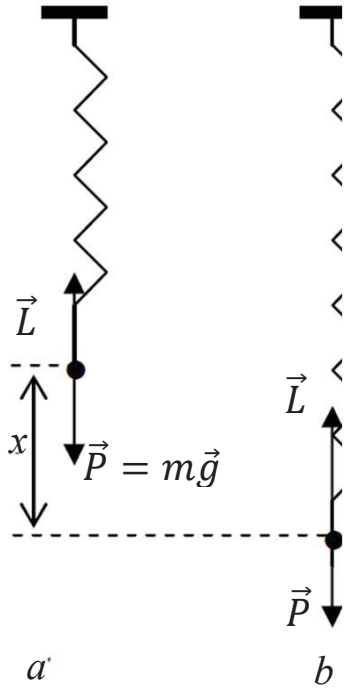
Problems for chapter 5

Problem 109. A conical pendulum casts a shadow on a vertical plane. What is the speed of the shadow at the point removed from the equilibrium position by distance a if it is known that the radius r of the circumference described by the weight of the pendulum is much smaller than the length of its thread l (see the figure).



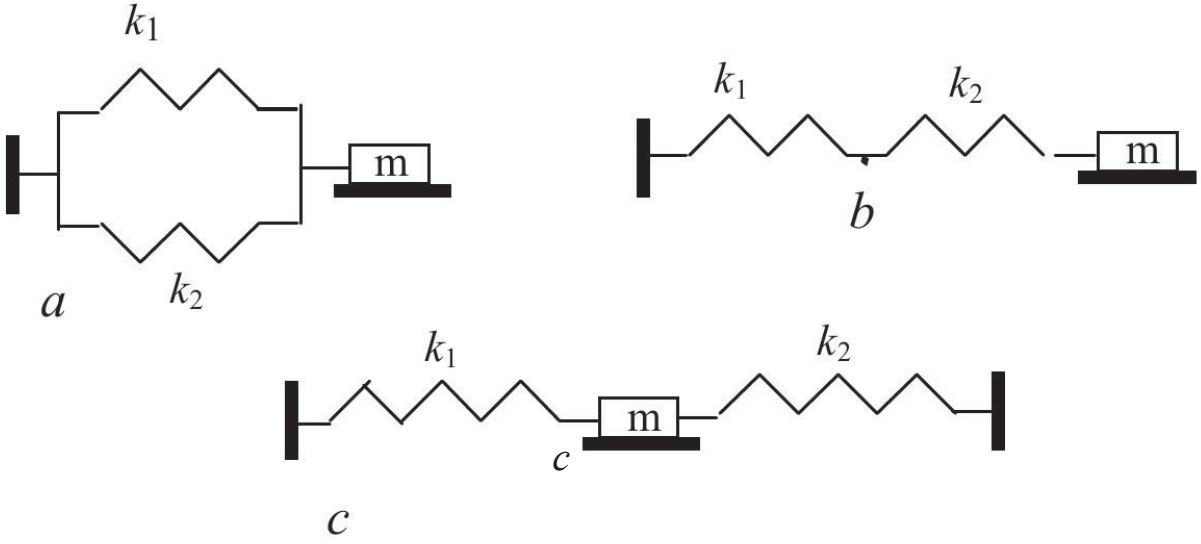
Problem 110. A steel ball is dropped without initial velocity onto a horizontal steel plate from height h. If we neglect the loss of energy of the ball during its motion, it will make oscillations similar to those made by a mathematical pendulum. What should be the length l of the string of a mathematical pendulum so that it oscillates with the period of an oscillating ball?

Problem 111. A weight of 0.1 kg is suspended on a spiral spring with a stiffness factor *k = 10 N/m* (see Figures *a and b*).



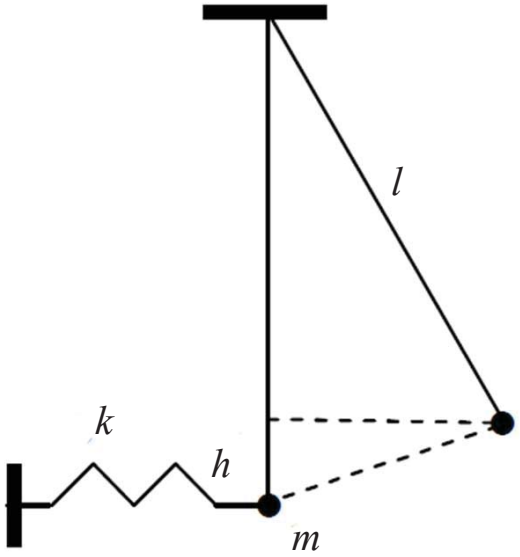
What forces act on the weight in the equilibrium position and also in the case of its displacement from the equilibrium position vertically downwards by the value of *x*? Which motion does the weight make, in cases *a and b*? How will its law of motion depend on the magnitude of the displacement? Determine the law of motion of the weight at a given deviation, if it is known that the maximum possible deviation of the weight from the equilibrium position is $4\ 10^{-2}$ m, and its motion began at a time equal to zero. Determine the kinetic $\mathcal{E}_k$ and potential $\mathcal{E}_p$ energy of the weight at a time of *0.0785 s*? The mass of the spring is negligible.

Problem 112. Calculate the oscillation periods of the systems shown in Figures a, b, c if the stiffness coefficients of the elastic elements and the mass of the load are known.

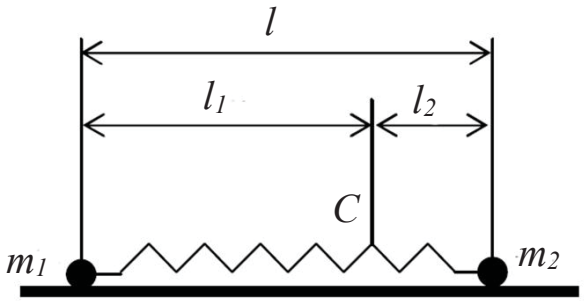


Problem 113. A mathematical pendulum, the weight of which is equal to m, and the length of the thread l, forms an oscillating system with the help of a connection made by an elastic element with stiffness k, one end of which is fixed motionless, and the other one is connected to the weight (see figure). Calculate the period of small oscillations of the

system, assuming that the mass of the elastic element is zero, and friction and air resistance can be neglected.



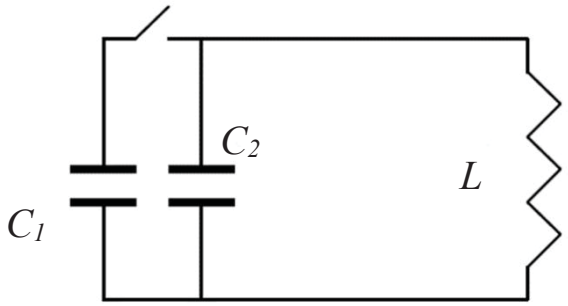
Problem 114. An oscillating system consisting of two masses m_1 *and* m_2, connected by an elastic element, is located on a smooth horizontal surface (see figure). The stiffness of the elastic element is k. Determine the period T of free oscillations of the system, deduced from the equilibrium position by deflecting the weights relative to its center of mass, if the weights during oscillations move along one straight line without friction.



Problem 115. The oscillating circuit consists of a coil connected in series with inductance $L = 8\ \mu H$ and a battery of two charged capacitors with the same capacity $C_1 = 10\ nF$. Determine the frequencies ω *and* ν,

as well as the period T of the free oscillations of the current $i(t)$ that occur when the circuit is closed for the cases of capacitors connected in series and in parallel. Losses in the circuit can be neglected.

Problem 116. A copper or ferrite core is introduced into the inductance coil of an oscillating circuit with inductance L connected to a constant capacitance C. How does the natural frequency of the circuit change?



Problem 117. In the oscillating circuit, capacitor C_1 is charged to voltage U, and capacitor C_2 is not charged (see figure). The capacitances of the capacitors are equal to each other and equal to C. Ignoring the resistance of the coil L, find the amplitude of the alternating current I_m arising in the coil conductor after the key is closed.

Solving the problems of chapter 5

109. A conic pendulum is a mathematical pendulum whose weight, suspended on a thread, forms a constant angle φ with the vertical and describes a circle in the horizontal plane. As can be seen from the figure of the problem condition, the magnitude x of the deviation of the shadow of the pendulum from the equilibrium position is

$$x = r \sin \varphi.$$

Since $r << l$, the shadow of the pendulum performs harmonic oscillations with a cyclic frequency ω. At the same time $\varphi = \omega t$, a $\omega = 2\pi/T$. The period of elastic oscillations T is determined by the formula

$$T = 2\pi \sqrt{\frac{1}{g}}$$

Speed of shadow movement

$$v_x = \frac{dx}{dt} = r\omega \cos \omega t\,; \;\; \cos \omega t = \frac{\sqrt{r^2 - a^2}}{r}\,; \;\; v_x = \sqrt{\frac{g(r^2 - a^2)}{l}}$$

110. $l = 2\frac{h}{\pi^2}$

111. Both in the first and in the second cases the force of gravity directed vertically downward and the force of spring tension applied to the weight according to Newton's third law vertically upwards act on the weight (see the figure of the problem condition). In the first case (Fig. *a*), these forces balance each other, and the weight is stationary relative to the Earth. In the second case (Fig. *b*), the force stretching the spring exceeds its tensioning force, and the weight moves downward with a deceleration. It is known that for relatively small deflections, the spring tension force (tension or compression) is proportional to the deflection, directed against it, and leads to harmonic oscillations of the weight. The law of motion in this case is described by a differential equation, the solution of which has the form

$$\mathrm{x} = A \sin\,(\omega t + \varphi_0),$$

where

A is the amplitude of oscillations,

ω is the cyclic frequency;

ωt is the phase of oscillations;

φ_0 is the initial phase.

Cyclic frequency and oscillation period T are related by the equiation

$$\omega = \frac{2\pi}{T}$$

In our case

$$\varphi_0 = 0,$$

$$T = 2\pi\sqrt{\frac{m}{k}};\ \ \mathcal{E}_k = \frac{1}{2}mv^2;\ \ \mathcal{E}_n = \frac{1}{2}mv^2$$

where $v = \frac{dx}{dt}$ is the speed of the weight;

T = 0,628 s;

t = T/8.

After appropriate substitutions we obtain

$$\mathcal{E}_k - \mathcal{E}_n = k\frac{A^2}{4} = 40\ J.$$

112. In the systems in the figures of the problem conditions the elastic elements are connected in parallel (*a, c*) and in series (*b*). The stiffnesses of systems with different connection of their elastic elements are found from the well-known relation

$$F = kx.$$

In the case of parallel connection of elastic elements:

$$x1 = x2 = x,\ F = F1 + F2\ and\ k = k1 + k2.$$

In the case of their connection in series:

$$x = x_1 + x_2, \qquad F = F_1 = F_2\ and\ k = \frac{k_1 k_2}{k_1 + k_2}.$$

The period of oscillation of the corresponding system is calculated by the formula derived from the equation of motion

$$T = 2\pi\sqrt{\frac{m}{k}}$$

113. Let the deviation of the system from the equilibrium position be expressed by the generalized coordinate *q*. In this case, the rate of change in the magnitude of the deviation of the system is $\frac{dq}{dt}$. There is no change in the total energy of the system, according to the law of conservation of energy, i.e. $\frac{d\mathcal{E}}{dt} = 0$. The total mechanical energy of the system in the absence of friction and air resistance at each moment of time is the sum of its kinetic energy $\mathcal{E}_k$ proportional to the square of

velocity $\frac{dq}{dt}$, and the potential energy $\mathcal{E}_p$ proportional to the square of deviation q

$$\mathcal{E} = Aq^2 + b\left(\frac{dq}{dt}\right)^2$$

It follows that

$$\frac{d\mathcal{E}}{dt} = 2\frac{dq}{dt}\left(Aq + B\frac{d^2q}{dt^2}\right) = 0$$

Since in dynamics $\frac{dq}{dt} \neq 0$, from the latter expression we obtain

$$\frac{d^2q}{dt^2} = -\frac{A}{B}q$$

The resulting equation is the equation of harmonic (undamped) oscillations (see Section 5.1) with cyclic natural frequency

$$\omega = \sqrt{\frac{A}{B}}$$

For harmonic oscillations of the system, as is known,

$$q = C\sin\omega t\,;\ \frac{dq}{dt} = C\omega\cos\omega t$$

The potential energy of the system in this case consists of the potential energy of the elastic element $kq^2/2$ and the potential energy of the weight mgh (see the figure of the problem condition).

Here

$$h = l - \sqrt{l^2 - q^2} \approx \frac{q^2}{2l}$$

(The approximate value of h is found by decomposing the radical into a power series and discarding all its terms above the second order of smallness, Appendix 2).

The kinetic energy of the system is equal to the kinetic energy of the weight

$$\frac{1}{2}m\left(\frac{dq}{dt}\right)^2$$

Comparing the original expression for total energy with its value for harmonic oscillations, we find that

$$A = \frac{1}{2}\sqrt{k + \frac{mg}{l}}; \text{ a } B = \frac{m}{2}$$

It is, in this way,

$$Y = 2\pi\sqrt{\frac{B}{A}} = 2\pi\sqrt{\frac{ml}{kl + mg}}$$

114. The oscillations of the weights (see the figure of the problem condition) occur with respect to the center of mass C, which is stationary with respect to them. The oscillations of the weights are harmonic and occur in antiphase (the instantaneous velocities are opposite in direction). The periods of oscillation equal to each other in connection with the immobility of the center of mass relative to the parts of the system are found by the usual formulas for harmonic oscillations. The stiffness coefficients k_1 *and* k_2 of parts l_1 *and* l_2 of the elastic element are respectively inversely proportional to their lengths so that

$$k_1 = k\frac{l}{l_1};\ \ k_2 = k\frac{l}{l_2}$$

The parts of the system l_1 *and* l_2 are inversely proportional to the masses of the weights.

$$l_1 = \frac{m_2 l}{m_1 + m_2} \text{ и } l_2 = \frac{m_1 l}{m_1 + m_2}$$

Substituting the data and the given values of masses and stiffnesses into the formula for the period of oscillation of the parts of the system gives

$$T = T_1 + T_2 = 2\pi\sqrt{\frac{m_1 m_2}{k(m_1 + m_2)}}$$

115. The equation of free oscillations for the generalized coordinate q occurring in an oscillating circuit in the absence of losses, as in the case of any other oscillating system, is

$$\frac{d^2q}{dt^2} + \omega^2 q = 0$$

If we choose the electric charge as the quantity q, then for the cyclic frequency of oscillations and it takes place

$$\omega^2 = \frac{1}{LC}$$

When the capacitors are connected in parallel, the capacitance of the battery, $C = 2C_1$ and when connected in series, $C = C_1/2$.

Substituting the given values gives:

for the parallel connection of capacitors $v = 400$ кГц, $\omega = 2{,}512$ МГц, $T = 2{,}5$ μs, and for their connection in series -$v = 800$ кГц, $\omega = 5{,}024$ МГц, $T = 1{,}26$ мкс.

116. Variable magnetic field of electromagnetic oscillations of the circuit induces inductive emf in the coil core according to the law of electromagnetic induction. Under its influence in the copper core, which is a conductor, vortex induction eddy currents arise, which, according to the Lenz law, weaken the coil's own magnetic field and, accordingly, its inductance. No currents occur in the ferrite core, which is a dielectric. On the other hand, a ferrite core made of ferromagnetic material increases the inductance of the coil many times over. Since, regardless of the type, all harmonic oscillations are described by the same equation, see solution of Problem 113, then

$$\frac{d^2q}{dt^2} = -\omega^2 q$$

At the same time

$$\omega = \sqrt{\frac{1}{LC}}$$

The natural frequency of the circuit will increase if a copper core is introduced into the inductor coil. If a ferrite core is introduced, the frequency will decrease.

117. Before the circuit is closed (see the picture of the problem condition) all energy of the system is contained in capacitor *1* and is equal by definition to $CU^2/2$. At the moment the circuit is closed, the current in the coil and therefore the electrical energy in it is zero, and the energy of the system is still concentrated in capacitors 1 and 2. At the same moment the initial charge CU is redistributed equally between the capacitors. According to the law of conservation of charge in the capacitor system according to the relation $CU = 2CU_0$, the voltage U_0 is set. If the capacitors are connected in parallel, their capacitances are added, so $U_0 = U/2$. The energy of the system decreases to $CU^2/4$.

By the time the current in the inductor coil reaches its maximum amplitude value I_m, the capacitors are discharged, all energy of the system is concentrated in the inductor coil and is equal by definition to $LI_m^2/2$. According to the law of conservation of energy $LI_m^2/2 = CU^2/4$. From here

$$I_m = U\sqrt{\frac{C}{2L}}$$

Problems for chapter 6

Problem 118. An electromagnetic wave of length $\lambda = 300\ m$ is modulated by a sound wave with frequency $\nu = 500\ Hz$. How many electromagnetic oscillations n fit in a sound wave in one period.

Problem 119. How will the wavelength and its frequency change when moving from one medium to another?

Problem 120. Water is slowly poured into a cylindrical tube of height $h = 1$ m. At what heights of water levels in the pipe will the sound of a tuning fork with a frequency of $\nu = 331\ Hz$ above it increase significantly? The speed of sound in air is *331 m/s*.

Problem 121. How will the tone of a tuning fork, wind instruments, and the human voice sound different in a decompression chamber compared to their tone in air?

Problem 122. A supersonic jet flew over the observer at a height of $h = 5\ km$ with a speed υ, twice the speed of sound u. How far away was the plane from the observer when he heard the sound?

Problem 123. A racing car with a siren turned on, sounding at a frequency of $\nu_0 = 400\ Hz$, raced towards a traffic policeman at a speed of 85 m/s. The policeman decided that the driver of the car violated the standard frequency of the siren (400 ± 25 Hz). So he demanded, using a whistle that sounded at a frequency of 15 kHz, that the car stop, but the driver did not respond to this command and drove by. Later, when the incident was investigated, the driver claimed that he had not violated anything, because his siren sounded at 400 Hz (as proof, he presented a frequency meter reading), and that he had not heard the policeman's whistle. The policeman, on the other hand, showed his whistle, claiming that it was impossible not to hear the sound of his whistle. The police officer who handled the case, to the surprise of those present, took out a notebook and pen and began to count something there. After a few minutes, he smiled and said that both the driver and the policeman were right, and he let the driver go in peace. Why did the police officer do that?

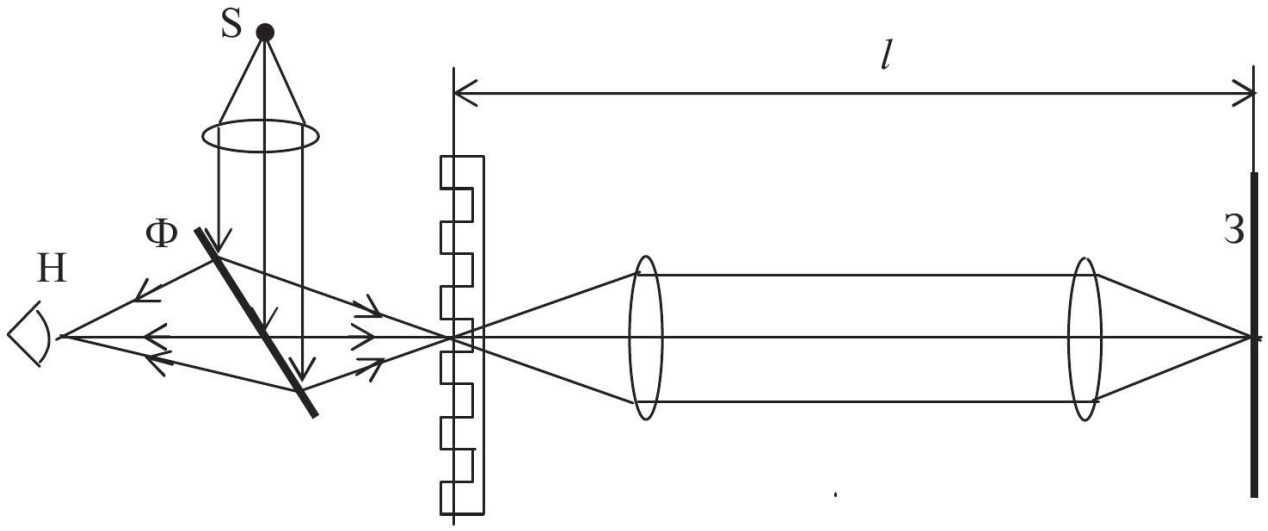
Problem 124. At what maximum distance l_{max} can a ship's radar detect a target at sea, if its antenna is at $h = 25$ m above sea level? At what frequency f should the radar emit pulses? How can the range of the radar be increased?

Problem 125. What is the de Broglie wavelength of the neutrino, as well as the proton, accelerated by the potential difference of U = 3,0 MV?

Problem 126. One of the first experiments that made it possible to measure the speed of light with a high enough accuracy was the Fizeau experiment. It consisted in that the light beam from the source S was passed with the help of a focusing device Φ through a narrow slit between the teeth of a rotating wheel K with the number of teeth $n = 720$ and was directed to the mirror 3, located at a distance $l = 8.7$ km (see the figure).

S

The light reflected off the mirror in the opposite direction. Once again passing through the slot between the teeth of the wheel, it was caught by the observer *H*. Calculate after Fizeau the speed of light with, if according to Fizeau the minimum speed of the wheel *f*, at which the light reflected from the mirror was not yet visible to the observer, was equal to 14.5 s^{-1}. What other more accurate methods of measuring the speed of light do you know

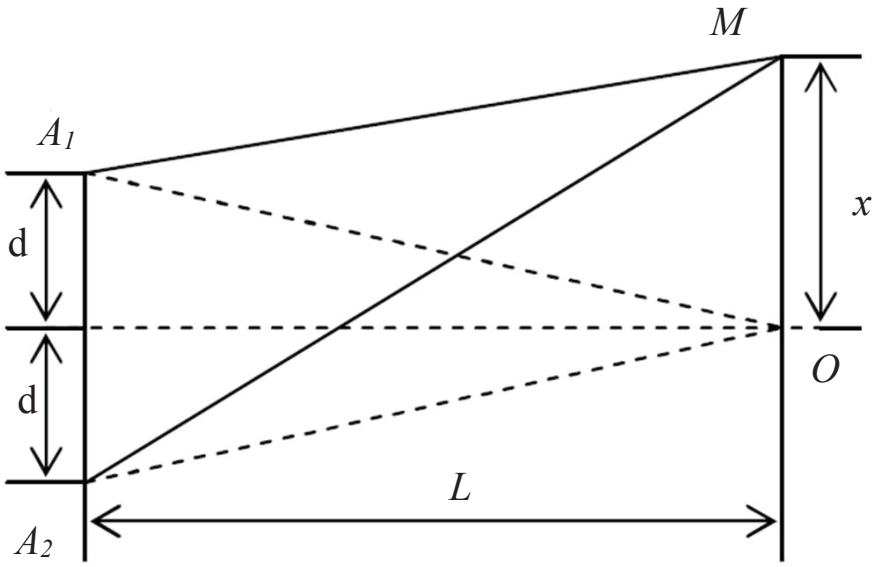


Problem 127. What explains the blue color of the sky? What color is the sky on the moon and why? How does white lettering on a red background look when illuminated with green light? How do you explain the red color of the setting sun and the evening dawn?

Problem 128. In interference there is a mutual weakening or amplification of the intensity of light rays in a certain area. Doesn't this violate the law of conservation of energy?

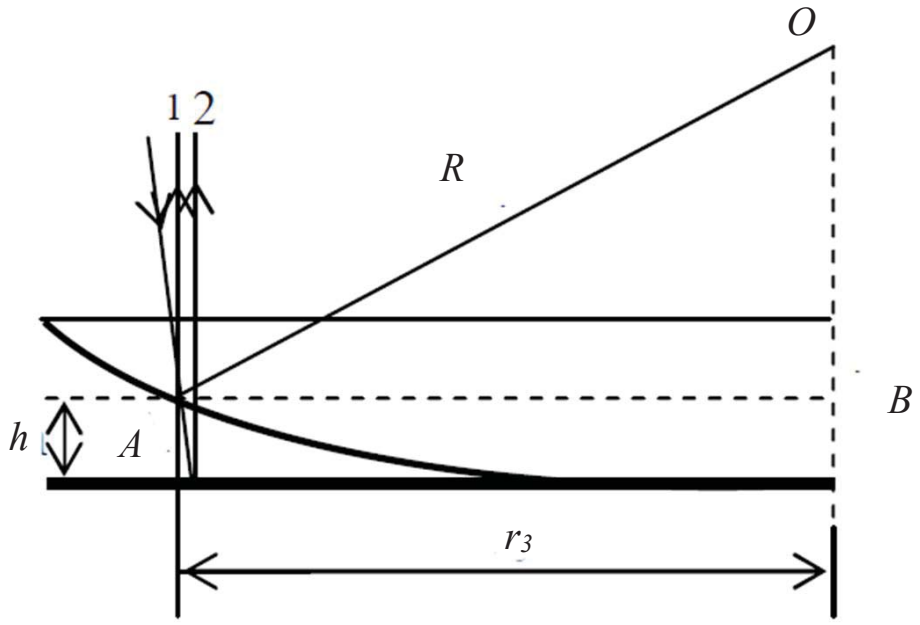
Problem 129. Why is the interference pattern observed only in sufficiently thin films (e.g., on fat films spread on the surface of solids, on films of gasoline on the ground, etc.)? The coloring of thin films resulting from interference is called iridescence. Are the colors reflecting the interference pattern spectrally pure?

Will there always be a maximum illuminance on the screen at point *O* (see figure), from two coherent monochromatic light sources A_1 *and* A_2? What is the distance *x (OM)* between the two nearest maxima (minima) of illumination if $A_1A_2 = 1$ *mm,* $L = 3$ *m,* $\lambda = 600$ *nm?*



Problem 130. A system consisting of a flat-convex lens, which lies convex side down on a flat horizontal glass plate, illuminated by a vertical beam of monochromatic light. When this system is observed in reflected light (rays 1,2), a dark circular spot surrounded by light and dark concentric rings can be seen. This picture is called Newton's rings (see figure).

Explain the appearance of Newton rings and calculate the radius r_3 of the third dark ring. Radius of curvature of convex side of lens $R = 1m$. The wavelength of light illuminating the system is $\lambda = 500\ nm$.



The deviation of the illuminating light beam from the vertical can be neglected.

Problem 131. A monochromatic beam of light with wavelength λ is incident at angle β on a diffraction grating with period d. It is known that for a normal incidence of light on a diffraction grating ($\beta = 0$) its formula is

$$d \sin \varphi = k\lambda$$

where φ is the angle between the direction of the beam reflected from the grating and the normal to the grating plane.

How will the diffraction grating formula change when $\beta \neq 0$?

Solving the problems of chapter 6

Determine the period of sound oscillations by the formula $\mathrm{T}_{sound} = 1/v = 2\,ms$. Electromagnetic wave propagation period

$$T_{em} = \frac{\lambda}{\mathrm{c}}, \qquad T_{em} = 10^{-6} s.$$

It is, in this way,

$$n = \frac{2 \cdot 10^{-6}}{10^{-6}} = 2000.$$

119. A wave of a given frequency v incident on the boundary of two media causes particles of the boundary surface to oscillate with the same frequency. These oscillations, in turn, are transmitted from surface particles, as sources of secondary waves, to all particles of the second medium and, therefore, cause a wave of the same frequency to propagate in it. This means that the frequency of the wave does not change when crossing the boundary of two or more media. However, the speed at which the wave propagates depends on the nature of the medium in which it propagates. Since wavelength λ is proportional to velocity, $\lambda = v/v$, the wavelength will change as much as its velocity by the same factor.

120. Under the action of the running sound wave emitted by the oscillating tuning fork and the wave reflected from the water surface of a

given level, a standing wave occurs in the air column of the tube. The node of this wave is formed at the surface of the water, and the antinode is formed at the open edge of the pipe. In a standing wave, as we know, the distance between the closest nodes is half the wavelength λ, and the distance between the node and the closest antinode is a quarter of the wavelength. It follows that the height of the air column

$$h_1 = \frac{\lambda}{4} + k\frac{\lambda}{2}$$

where *k* is the number of possible wave nodes, k =0; 1; 2;...

Since $\lambda = \frac{v}{v} = 1$ м, and the height of the tube is also 1 m, the value of *k* can only take values 0 and 1. At *k* =0 h_1 = 0.25m, and the water level is -0,75m; at *k* = 1 h_1 = 0.75m, and the water level is 0.25m.

121. The tone (pitch) of a monochromatic sound wave emitted by a body that makes harmonic oscillations, such as a tuning fork, is determined by its frequency v and the properties of the sounding body. Musical instruments, vocal cords, etc. emit sound, however, in a certain spectrum, which is a mixture of different frequencies (harmonics). The tone of the sound is determined by the frequency of the harmonic with the greatest intensity (amplitude), and the number of harmonics determines the timbre of the sound.

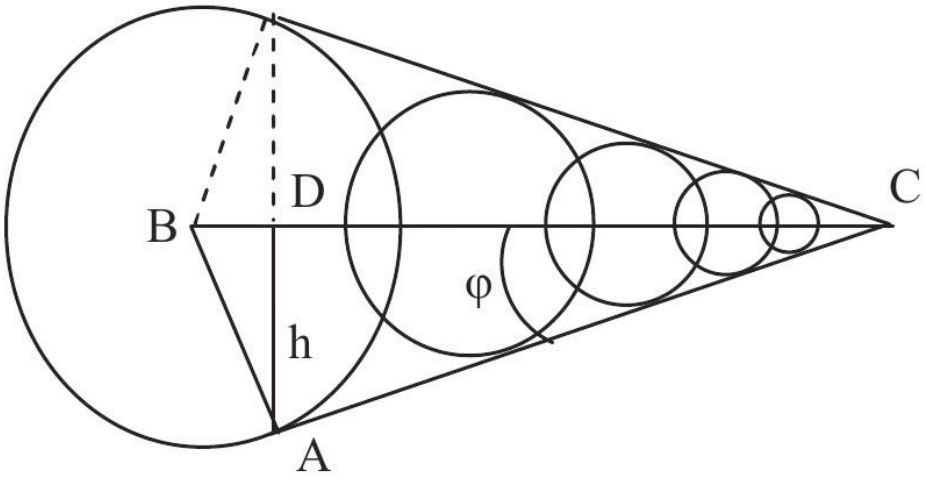
The human larynx, as well as the trumpet of wind musical instruments, are resonators of sound oscillations (acoustic resonators). A resonator, as we know, is a device that amplifies sound waves whose frequencies coincide with its natural frequency, calculated according to the formula

$$v = \frac{v}{\lambda}$$

Unlike the frequency and length of a monochromatic sound wave, the propagation velocity of the harmonics of complex sound depends significantly on the propagation medium. In decompression chambers, air, which is basically a mixture of oxygen and nitrogen, is replaced by a breathing gas mixture consisting of oxygen and helium. As a result, the average molar mass of the medium is significantly reduced in them and,

accordingly, the speed of sound and the frequency of natural vibrations of resonators (larynx, wind instrument pipes, etc.) are increased. Increasing their resonating natural frequency leads to an increase in tone and loudness.

122. Let at the moment of time when the sound originated at point *B* has reached the observer at point *A*, the plane is at point *C* (see figure). This means that during the time t, between the moment of sound emission and the moment when the observer heard it, the plane traveled the distance $BC = vt$. During the same time the sound traveled the distance to the observer $BA = ut$. The front of the sound wave arising in the air during the flight of an plane is an envelope of spherical waves emitted at each of the previous moments of time. It forms a cone, which is called a sonic cone. In other words, *AB* is the radius R of the spherical front of the wave emitted at point *B*. From right-angled triangle *ABC* it follows that $AB = BC \sin\varphi$, and from triangle *ADC,* $\sin\varphi = \frac{h}{AC}$. Therefore $\frac{AC}{h} = \frac{BC}{AB} = \frac{v}{u}$ and $AC = 10\ km.$

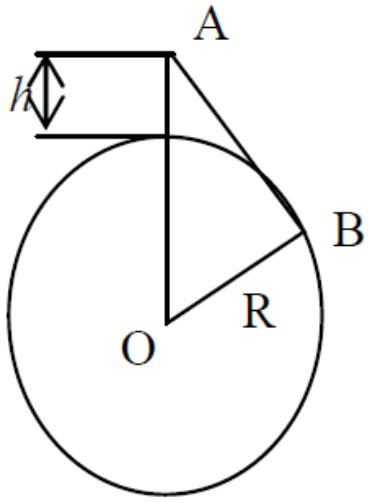


123. The police officer did the right thing, because he, unlike the policeman and the driver, knew physics and remembered that the Doppler effect applies in acoustics, according to which the frequency of acoustic waves perceived by the observer depends on the relative speed of the sound source and the observer. In the first case it was

$$v = v_0 \sqrt{\frac{1 + \frac{u}{v_{sound}}}{1 - \frac{u}{v_{sound}}}}$$

Thus, the police officer heard the siren sounding at a frequency of 526 Hz, while to the driver the whistle sounded at a frequency of 19.74 kHz, which is usually already poorly perceived by the human ear.

124. Let the upper point *A* of the radar antenna, which radiates electromagnetic waves, be at a point at height *h* above sea level (see figure). It is known that waves, including electromagnetic ones, propagate in a straight line along the radial lines of the sphere. On the other hand, one of the most important properties of wave processes is diffraction, which consists in the fact that waves bend around obstacles on their way, the size of which is comparable with the wavelength. However, in the ultrashort wavelength range in which radars operate, the deviation from straightness caused by diffraction is so sm $_A$ it can be neglected. In this case, the contact point *B* of the radar beam *AB* with the spherical sea surface of the globe centered at point O determines the desired maximum range $AB = l_{max}$ of target detection.



Since it follows from right-angled triangle ABO that ${l_{max}}^2 = (R + h)^2 - R^2$, then finally

$$l_{max} = \sqrt{h(2R + h)} \approx 18\ km$$

So that the pulses emitted by the radar do not overlap and interfere with each other, they must be emitted at a certain interval, so that the previous pulse reflected from the target returns to the locator receiver before the next pulse appears. This means that the period of emission *T* must satisfy the inequality

$$T \geq \frac{2l_{max}}{c}$$

and

$$f \leq \frac{c}{2l_{max}} = 3800\ s^{-1}$$

The radar range is determined not only and not so much by the maximum possible distance of the locator, but mainly by its power *N*. Since the front of the emitted wave is a sphere, the amount of emitted energy in each given point of space, is distributed uniformly inversely proportional to the square of the sphere radius R, equal to the distance from the radar to the target. Otherwise, the intensity I (power per unit surface of the wave front)

$$I \approx \frac{N}{R^2}$$

On the other hand, the target, reflecting the radio signal, is a source of secondary waves. Therefore, the intensity of the signal received by the radar is, in turn, inversely proportional to the square of its distance to the target. Thus, the resulting intensity is inversely proportional to the fourth power of the distance to the target.

Consequently, to increase the range of target detection, for example by 4 times, you need to increase the power of the locator by $4^4 = 256$ times.

125. The de Broglie wavelength is calculated by the formula

$$\lambda = \frac{h}{m_0 v}$$

The speed of motion of neutrinos is known to be close to the speed of light in the vacuum, i.e. we can assume that $v_n = c = 3 \cdot 10^8\ m/s$.

The assumed rest mass of the neutrino $m_0 \approx 30\,eV/s^2 = 5{,}3 \cdot 10^{-35}\ kg$. It is, in this way,

$$\lambda_n = 4{,}2 \cdot 10^{-4} \mu m$$

This wavelength is within the range of X-rays.

For the proton $eU = m_p v^2/2$. From here

$$\lambda_n = \frac{h}{\sqrt{2eUm_p}} = 1{,}7 \cdot 10^{-14}\ m$$

This wavelength is in the range of hard gamma rays.

126. If during the time *t*, while the light travels to and from the mirror, the wheel manages to turn by such a minimum angle φ that a tooth takes the place of the slit, the light is not picked up by the observer. The observer will see the light, if another nearby slit takes the place of this slit in the specified time.

According to the conditions of the problem

$$\varphi_{min} = \frac{2\pi}{2n} \text{. a } f_{min} = \frac{\varphi_{min}}{2\pi t}$$

where $t = \frac{2l}{c}, therefore\ c = 4f_{min}nl \approx 3{,}14 \cdot 10^{84}\ m/s$.

French physicist Léon Foucault used a rotating mirror instead of a rotating wheel, which allowed him to obtain a more accurate value of the speed of light c = 2.98 10^8 m/s. Even more accurate value was obtained by Albert A. Michelson using an improved method of rotating mirrors and the interferometer he invented. He managed to get a very close to the modern value of c = 2.99796 10^8 m/s.

127. The Earth's atmosphere has a large scattering effect. The centers of scattering are molecules of atmospheric gases and dust particles of negligible size. As the English physicist John William Strutt, 3rd Baron Rayleigh, showed, the intensity of light scattering on microparticles, the size of which is small compared to the wavelength of light, is inversely proportional to the fourth power of the wavelength. From the modern point of view scattering has a quantum-mechanical inelastic character. It is considered as a result of absorption by

microparticles of photons of one frequency with their subsequent excitation and emission by them of photons of another frequency of higher energy of the short-wave spectrum. As a result, when white sunlight passes through the atmosphere, photons of the long-wave spectrum, that is, red light and light of red hues are absorbed, and light of blue hues is emitted, taking on average a blue coloring. We see the Sun itself as yellowish, because some of the short-wave light scattered by the atmosphere does not reach us

At sunrise and sunset, when the Sun is inclined toward the horizon, its rays travel the greatest distance in the atmosphere, which causes scattering, in addition to blue and blue rays, also yellow and green. In this case, mainly red rays come to us from the Sun. These same rays, especially when the air is saturated with dust particles or droplets of moisture, color the sky and the clouds floating in it near the rising or setting Sun red, creating morning and evening dawn. The same is observed at moonrise and moonset.

On the Moon, where the atmosphere and, consequently, the centers of scattering and radiation are absent, the sky is black, on which clearly stands out the white-hot Sun, the blue Earth, whose atmosphere secondarily emits blue scattered sunlight, and other luminaries.

A white lettering on a red background reflects the entire spectrum of visible light and looks white, but when illuminated by green light it will reflect green rays, which, mixed with the red rays of the red background, will give the impression of a dark coloring.

128. No, it is not violated, because at interference the mutual damping of waves in some areas is exactly compensated by their mutual amplification in other areas. Energy does not disappear, but is only redistributed in space.

129. If the film thickness exceeds the wavelength of incident visible light (tenths of a micron), then the waves reflected from the plane-parallel surfaces of the film are not strictly coherent. This is because the coherence of the rays is lost as they move inside the film due to interaction with incoherent rays falling on the film. Moreover, even if approximate coherence is preserved, the density of maxima and minima

increases sharply with increasing film thickness and the latter become indistinguishable.

The rainbow colors produced in thin film by interference are not purely spectral, as they result from the attenuation of some spectral components and amplification of others and their subtraction from the white color.

At a point distant from the coherent sources, both maximum illuminance in the case of in-phase sources and minimum illuminance in the case of antiphase sources can occur.

The required distance x will be found from the condition of maxima (minima) for $k = 1$, according to which the difference of path equals the wavelength. This condition, as can be seen from the figure of the problem condition, has the form

$$\sqrt{L^2 + (x+d)^2} - \sqrt{L^2 + (x-d)^2} = \lambda;$$

$$d = \frac{A_1 A_2}{2}$$

Substitution yields

$$x = 1{,}27 \text{ mm}.$$

130. Newton rings appear as a result of interference resulting from the superposition of coherent light beams reflected from the surface of a lens and a flat plate forming a thin air layer (thin plate). The condition of the interference minimum (dark ring) is determined by the difference in travel d and has the form

$$d = (2k+1)\frac{\lambda}{2}$$

On the other hand, the difference in travel, taking into account the change in beam travel by $\lambda/2$ in its reflection, is $d = \lambda/2 + 2h$, so

$$2h = k\lambda.$$

Let us consider an approximately vertical beam of parallel rays incident at point A (see the picture of the problem condition) of the convex surface of the lens, which is at distance h from the flat plate. Draw

the radii R from this point to the center O of the lens, and from the point O perpendicular to the plate. Since $h << r_k$ and R, it follows from the drawing of the figure that $h = r_k^2/2R$. After appropriate substitution we obtain that

$$r_k = \sqrt{kR\lambda}$$

For the third ring ($k=3$)

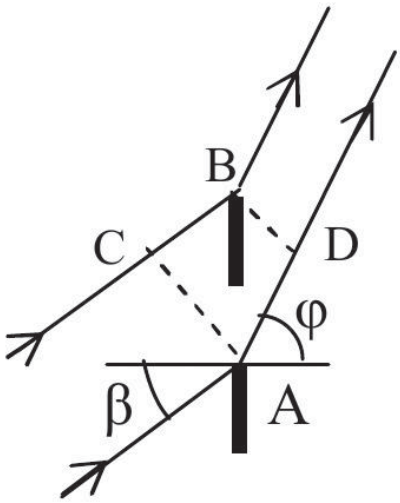
$$r_3 = \sqrt{3 \cdot 500 \cdot 10^{-9}} = 1{,}23\ mm$$

131. As can be seen from the figure, the difference of path between the rays reflected by the edges of neighboring slits is

$$AD - BC = d(\sin\varphi - \sin\beta).$$

From this we get the following formula for the diffraction grating

$$d(\sin\varphi - \sin\beta) = k\lambda$$



Israeli Independent Academy for the Advancement of Science

L. Preigerman

M. Brooke

MODERN PHYSICS COURSE

Under the editorship of

Doctor of Philosophy Professor O.E. Baksanskiy

in 5 parts

Third Edition

Part 5. Basics of quantum physics

Recommended by the Academic Council of the Russian State Classical Academy named after Maimonides (Rambam) for undergraduate and graduate students, and for secondary and high school physics teachers.

Moscow URSS 2014

Table of contents

Chapter 7. Fundamentals of Quantum Physics

Physics, as you know, has taken a very difficult, sometimes quite twisting and dramatic path of development. In ancient times, it was not yet an independent science, but was considered part of the philosophy studying Nature (Natural Philosophy), and had, like all philosophy, a purely speculative nature. Thinkers of this period, including Heraclitus, Democritus, Epicurus, Plato, Aristotle, and others, sought to construct a general conception of the world from which truths reflecting its particular manifestations would emerge. Most of them denied the significance of experience, which follows from sense perceptions that, in their view, distort the true picture of the world. The penetration into the essence of things by means of pure mind alone allowed them, nevertheless, to make a number of ingenious conjectures related to the ideas of matter and spirit, space and time, imaginary perceptions and ideas, atoms and matter, vibrations and waves, sound and light, motion and development, etc.

7.1. Pre-Quantum period in the development of physics.

However, the science of antiquity, detached from practice, proved to be unproductive and, therefore, already in the Middle Ages it gave way to a contemplative science based on observation and physical experiment. At its foundation stood such giants of medieval science as Leonardo da Vinci, Gilbert, Galileo, Kepler, Newton, Descartes, Huygens, Leibniz and others.

The new experimental science created by them, later called classical science, developed according to the level of experiment technique from simple to complex and was reduced to the theoretical generalization of the results of observations and experiments. Abstract comprehension on the basis of mathematical analysis of established experimental facts led science to the disclosure of the nature and mechanisms of the behavior of surrounding things and the establishment

of the rules of this behavior (laws), uniquely linking the state of things with their properties and the surrounding conditions.

The results of observations and physical experiments have increasingly strengthened scientists in the conviction that the patterns of functioning of real reality are strictly deterministic and objectively determined by its material nature. Hence it was concluded that the laws of Nature, established on the basis of observation and experiment, are universal and depend neither on the place and time of their observation nor on the observer himself. On this basis these laws were given the status of objective, absolute and eternal truths. In doing so, practice was adopted as the only criterion of truth, ruling out the possibility of any arbitrary interpretation. In other words, inferences, no matter how logical they might seem, could not be taken into account as evidence of truth unless they were corroborated by observation or experiment.

The above approach was in full accordance with direct perceptions of the surrounding reality and was confirmed by numerous projects implemented in practice, as well as the realization of predictions expressed in accordance with the scientifically established laws of Nature.

The triumph of the principle of determinism was the high accuracy of the French physicist Laplace's predictions of the behavior of the planets of the solar system, which were confirmed in practice. His fundamental work "Celestial Mechanics" is still used today to predict the timing of lunar and solar eclipses, the planetary closest approaches, the movements of comets, etc. The principle of determinism is also confirmed by the practice of daily activities and technical progress. As for the so-called random events that cannot be predicted, they were seen not as evidence of a violation of the principle of determinism, but as the result of the limitations of our knowledge. It was from this point of view that phenomena of Nature, including atmospheric cataclysms, tectonic processes acting in the lithosphere, such as volcanic eruptions,

earthquakes, etc., which were not yet fully understood, were considered. All of them were assumed to be regular, causally conditioned and considered as conditionally random. The probability assessing the possibility of their occurrence was characterized by a subjective value that increased with the level of knowledge.

Pre-quantum physics and its underlying postulates of determinism, causality and universality of Nature's processes for a long time, up to the beginning of the 20th century, were considered unshakable. However, facts gradually began to accumulate that entered into an irreconcilable contradiction with them.

7.2. Physics at a Crossroads

With the penetration of science into the microcosm and the depths of the cosmos, serious doubts have arisen about the possibility of obtaining reliable knowledge at all structural levels. Already during the transition to atomic-molecular levels, it became obvious that there is no real possibility to describe the thermal processes taking place there using only the laws of classical mechanics. This is due to the fundamental impossibility of experimentally obtaining all the necessary information for compiling the corresponding equations of motion of particles, as well as determining the initial conditions necessary for the complete solution of the mechanical problem. In order to keep the principle of determinism intact in this case, statistical physics was created, which began to consider macro-processes as the result of averaging of micro-processes on sufficiently large time intervals comparable with the time of observation. The theory of relativity, even more so than statistical physics, has introduced into scientific thinking the notion that our knowledge is relative. At the same time, the conviction of the unconditional objectivity of cognitive processes, based on the notion of their independence from the observer, was shaken. Such an approach, from the point of view of the theory of relativity, is purely approximate, since the observer himself

is part of the world he observes. The transition to the study of microsystems at the levels of the internal structure of atoms, atomic nuclei and the depths of space corresponding to the initial stages of the Universe has further shown that the same real systems under the same conditions behave differently so that their behavior can be predicted only with a certain probability.

It all began with the discovery of facts related to the interaction of matter and radiation, which did not fit into the laws of mechanics and electrodynamics. Many of these facts have already been mentioned above (see section 6.6.4). To resolve this crisis, some physicists have proposed hypotheses according to which matter and radiation under certain conditions lose their inherently specific corpuscular and wave properties. Radiation, along with wave properties, acquires corpuscular properties, and matter, along with corpuscular properties, acquires wave properties.

This means, as we already know, that between a free particle with discrete values of energy $\mathcal{E}$, momentum P, and a monochromatic wave with cyclic frequency ω and wave number k there is an unambiguous correspondence such that the same relations apply to both

$$\left.\begin{aligned} \mathcal{E} &= \hbar\omega \\ P &= \hbar k \end{aligned}\right\} \tag{7.1}$$

This hypothesis is known as wave-particle duality. It was the result of generalization of experimental data and allowed to coordinate the theory based on the laws of mechanics and electrodynamics, which did not contain the wave-particle duality, with practice, where it was unambiguously manifested. This hypothesis helped to explain the whole set of facts that did not fit into the theory. In order to better understand the peculiarities of quantum physics based on this hypothesis, it is necessary to consider the above-mentioned facts in more detail.

7.2.1. Thermal radiation

Thermal radiation is the electromagnetic radiation emitted by matter and arising from changes in its internal energy. Thermal radiation, from a classical point of view, has a continuous spectrum with intensity maxima at certain wavelengths. This radiation, as experience shows, takes place for any bodies in any state at any temperature.

Let us consider an arbitrary body and mentally surround it with a shell with a perfectly reflecting inner body surface so that the radiation reflected by the shell from the body is absorbed by the body again. As a result, there is a distribution and continuous exchange of energy between the body and the radiation that fills the space formed by the body and the shell (Fig. 7.1).

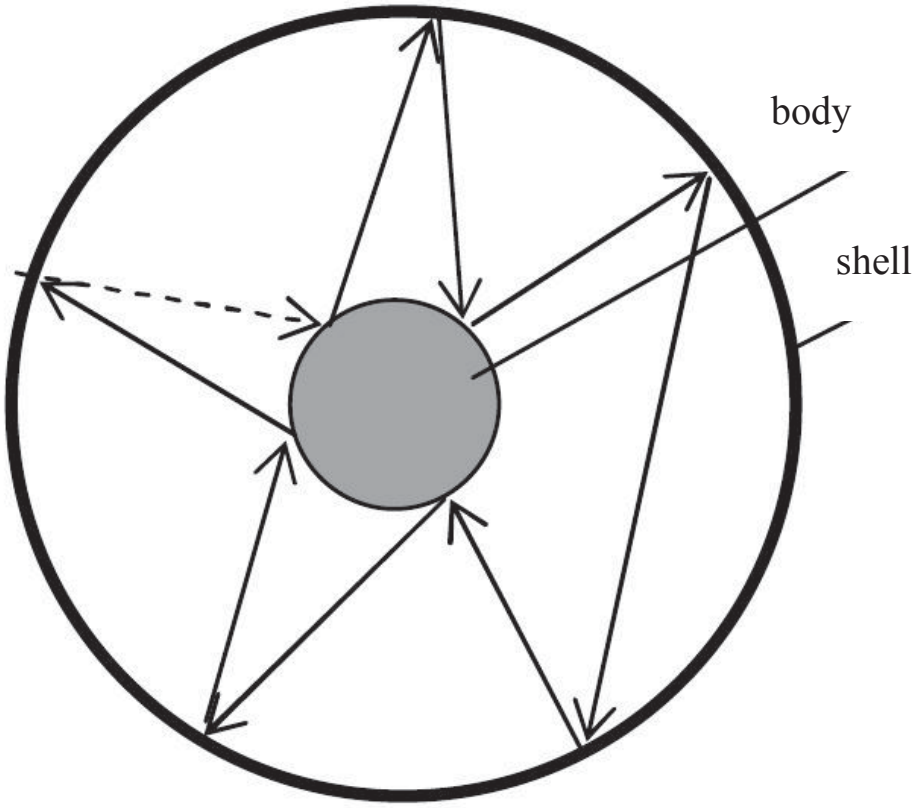


Figure 7.1.

If the distribution of energy between the body and the radiation remains constant for each wavelength of radiation, the state in the body-

radiation system is called equilibrium and is characterized by a certain temperature. As follows from experience, only thermal radiation can be equilibrium.

The amount of energy emitted by a unit of body surface in the entire wavelength spectrum per unit time in all directions is called the energy luminosity of thermal radiation R_T. The sign "$_T$" here and hereafter means that these quantities characterize radiation at temperature T.

Let the radiation flux dR be emitted in the frequency interval $d\omega$, with this $_{\omega,T}$ will be

$$dR_{\omega,T} = r_{\omega,T}\, d\omega \tag{7.2}$$

where $rR_{\omega,T}$ is the coefficient of proportionality called the spectral density of energy luminosity.

As can be seen from (7.2),

$$r_{\omega,T} = \frac{dR_{\omega,T}}{d\omega}$$

and

$$R_{\omega,T} = \int_0^\infty r_{\omega,T}\, d\omega \tag{7.3}$$

Let the energy flux $d\Phi_{\omega,T}$ fall on a unit of body surface dS in the frequency range $d\omega$ per unit time, so that part of this flux $d\Phi'_{\omega,T}$ is absorbed by the body and part $d\Phi''_{\omega,T}$ is reflected by it, then

$$d\Phi_{\omega,T} = d\Phi'_{\omega,T} + d\Phi''_{\omega,T} \tag{7.4}$$

The quantities designated by

$$a_{\omega,T} = \frac{d\Phi'_{\omega,T}}{d\Phi_{\omega,T}} \text{ and } b_{\omega,T} = \frac{d\Phi''_{\omega,T}}{d\Phi_{\omega,T}} \tag{7.5}$$

are called absorptivity and reflectivity, respectively.

From (7.4) and (7.5) we see that

$$a_{\omega,T} + b_{\omega,T} = 1 \tag{7.6}$$

If a body completely absorbs the radiation falling on it, then it is called **black body**. For it

$$a_{\omega,T} + b_{\omega,T} = 1 \tag{7.7}$$

The amount of thermal radiation energy per unit volume is called the **volumetric energy density** of thermal radiation U_T . For equilibrium radiation, the energy density is uniformly distributed over the volume and depends only on the temperature T.

The quantity

$$U_{\omega,T} = \frac{dU_T}{d\omega} \tag{7.8}$$

is called the **spectral energy density**. It follows from (7.8) that

$$U_T = \int_0^\infty U_{\omega,T} d\omega \tag{7.9}$$

Let inside a closed shell there are several bodies of different nature, which are in equilibrium with radiation at some temperature T. Since for equilibrium radiation, by definition, the internal energy remains constant, the body that emits more energy must absorb more of it. In other words, $r_{\omega,T}$ is proportional to $a_{\omega,T}$. This means that

$$\left(\frac{r_{\omega,T}}{a_{\omega,T}}\right)_1 = \left(\frac{r_{\omega,T}}{a_{\omega,T}}\right)_2 = \cdots = \left(\frac{r_{\omega,T}}{a_{\omega,T}}\right)_i = \cdots = f(\omega, T) \tag{7.10}$$

where 1, 2,3,... i, ... - are the numbers of the corresponding bodies.

The equation (7.10) is called Kirchhoff's law of thermal radiation, and the function $f(\omega, T)$ is called the **universal Kirchhoff function**. For a black body $a_{\omega,T} = 1$, so as follows from (7.10),

$$r_{\omega,T} = f(\omega, T) \tag{7.11}$$

The universal Kirchhoff function according to (7.11) is the emissivity of a black body.

Let us consider a small hole of area S_0 in the opaque wall of a large closed cavity S_N such that

$$\frac{S_0}{S_N} \ll 1 \qquad (7.12)$$

It follows from (7.12) that the probability of radiation entering the cavity returning to hole S_0 is practically zero (Fig. 7.2).

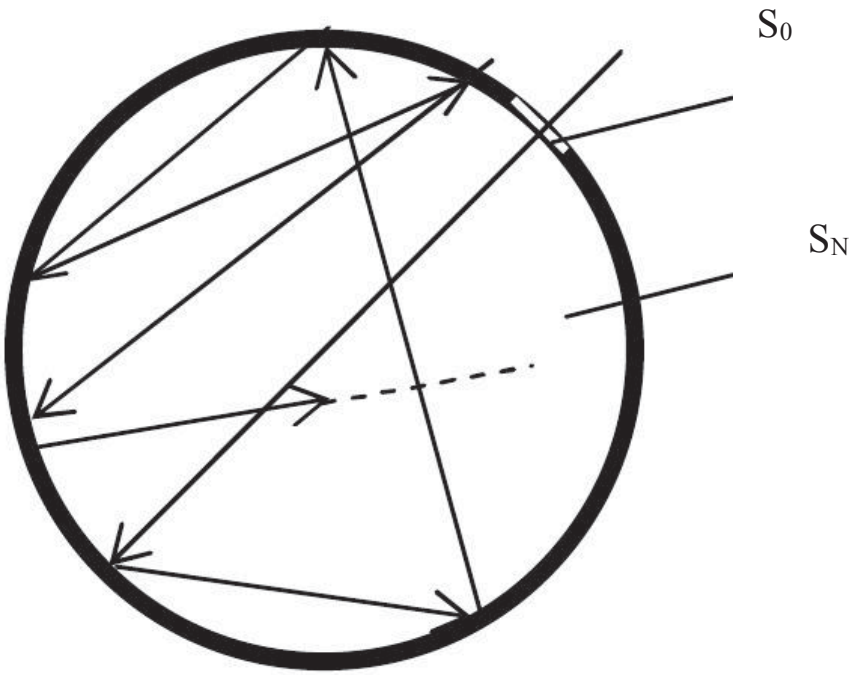


Figure 7.2.

All energy incident on hole S_0 is absorbed by cavity S_N, so so the following equation is true for this hole

$$a_{\omega,T} \approx 1.$$

It follows that a small hole in a large cavity is a perfect model of a black body. Since thermal radiation is inherent to any body, then according to Kirchhoff's law for this radiation at a given temperature T

$$r_{\omega,T} = f(\omega, T) \qquad (7.13)$$

By measuring $r_{\omega,T}$ of such a device, it is therefore possible to determine $f(\omega, T)$ for each given temperature. It follows from experience that the dependence of $f(\omega, T)$ on frequency and temperature has the form shown in Figure 7.3

Using the curves (Fig. 7.3), the value of the energy luminosity equal to the area covered by the corresponding curve is determined.

$$r_{\omega,T} = f(\omega, T),$$

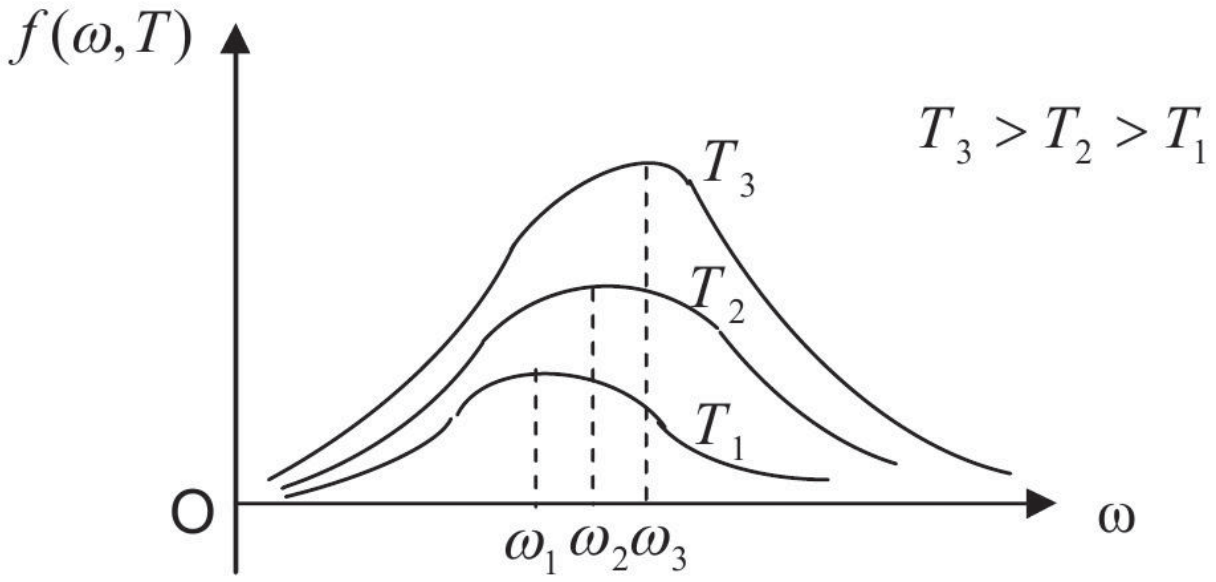


Figure 7.3.

Experimentally, i.e. using curves (see Figure 7.3), the Stefan-Boltzmann and Wien laws are established (see Section 6.6.2).

7.2.1.1. Classical theory of thermal radiation

The classical theory of thermal radiation was created by the English physicists John William Strutt, 3rd Baron Rayleigh and James Jeans in 1900-1909. Here are the main points of this theory. Let the cavity in the form of a rectangular box $l_1 \times l_2 \times l_3$ filled with thermal equilibrium radiation at temperature T. Let us assume that the walls of the cavity are completely impermeable and that the radiation in the cavity is in a state of equilibrium. We already know that the radiation inside such a cavity

can be represented as a set of standing waves (see Section 6.6.1). Let us first consider the one-dimensional version, which is modeled by a mechanical analogue in the form of a string of length *l* stretched between two points. Let's call the type or mode of oscillations of the n-th order a sinusoidal wave, which turns to zero at the ends of the string (standing wave nodes) and contains *n* half-waves. For the *n*-th order mode, we can thus write

$$\frac{l}{\lambda/_2} = n, \text{ n} = 1, 2, 3, \ldots \tag{7.14}$$

From (7.14) it follows that the wave number of the *n*-th order mode

$$k_{B_n} = n\frac{\pi}{l} \tag{7.15}$$

The value of the interval between two consecutive wave values *Δk* according to (7.15) is

$$\Delta k_B = (n+1)\frac{\pi}{l} - n\frac{\pi}{l} = \frac{\pi}{l}. \tag{7.16}$$

The number of modes *dN* per unit interval *Δk*, according to (7.16),

$$dN = \frac{\Delta k_B}{\Delta k_B} = \frac{l dk_B}{\pi} \tag{7.17}$$

Let us choose the xOyz coordinate system and consider a wave vector $\vec{k}_B$ with projections k_{Bx}, k_{By}, k_{Bz}, taking both positive and negative values. Such a vector describes not a standing wave, but a running wave. The number of waves running in one direction will obviously be two times less than that obtained in relation (7.17), so

$$dN(k_{B_i}) = \frac{l dk_{B_i}}{2\pi}, \text{ where } i = x, y, z \tag{7.18}$$

The number of waves in the whole cavity of the box is respectively

$$dN(k_B) = \frac{l_1 l_2 l_3}{(2\pi)^3} dk_{B_x} dk_{B_y} dk_{B_z} \tag{7.19}$$

or

$$dN(k_B) = \frac{V}{(2\pi)^3} d3k_B \tag{7.20}$$

where V is the volume of the box,

d^3k_B is an infinitesimal volume in the phase space of the wave vector $\vec{k}_B$ with axes k_{Bx}, k_{By}, k_{Bz}.

In general, for the considered system, the phase space consists of configuration space *(x, y, z)* and k_B -space (k_{Bx}, k_{By}, k_{Bz}). It is six-dimensional and is defined by dimensions *x, y, z,* k_{Bx}, k_{By}, k_{Bz} (see Section 2.5).

Since in this case the wave is spherical, the volume d^3k_B in k_B -space is the elementary volume of the spherical layer between k_B and $k_B + dk_B$ (Fig. 7.4).

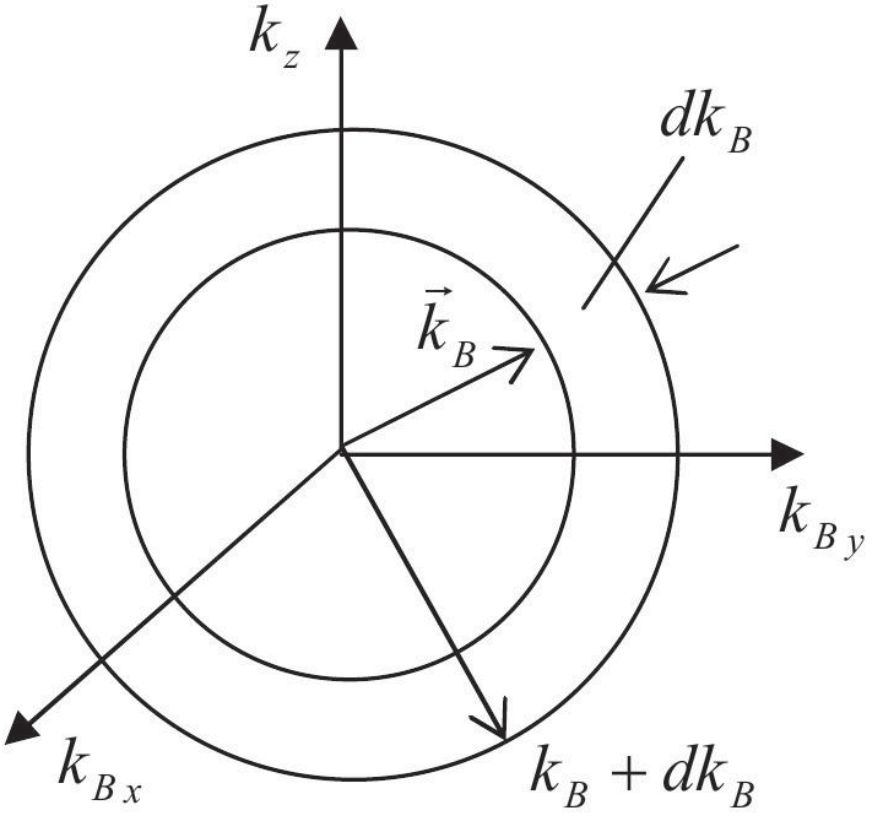


Figure 7.4.

It follows from Figure 7.4 that

$$d^3k_B = 4\pi k_B^2 dk_B \tag{7.21}$$

Substituting (7.21) into (7.20) gives

$$dN(k_B) = \frac{4\pi k_B^2 V dk_B}{(2\pi)^3} \tag{7.22}$$

The wave number of k_B radiation is, by definition, equal to

$$k_B = \frac{\omega}{c} \tag{7.23}$$

where c is the speed of light in a vacuum;

Therefore,

$$dN(\omega) = \frac{4\pi\omega V d\omega}{(2\pi)^3 c^3} \tag{7.24}$$

Due to the fact that we are talking about the natural oscillations of an electromagnetic wave, each wave vector $\vec{k}_B$ corresponds to two oscillations with mutually perpendicular polarizations, so finally

$$dN(\omega) = \frac{V\omega^2 d\omega}{\pi^2 c^3} \tag{7.25}$$

Concentration of waves

$$dn(\omega) = \frac{dN(\omega)}{V}$$

From here

$$dn(\omega) = \frac{\omega^2 d\omega}{\pi^2 c^3} \tag{7.26}$$

Assuming that the energy per mode is kT , we obtain

$$dU_{\omega,T_1} = kT \tag{7.27}$$

where k is Boltzmann number.

From here, taking into account (7.26), we find that the total energy

$$dU_T = \frac{\omega^2 kT}{\pi^2 c^3} d\omega \tag{7.28}$$

and the spectral density

$$dU_{\omega,T_1} = \frac{dU_T}{d\omega} = \frac{\omega^2}{\pi^2 c^3} kT \tag{7.29}$$

The relation (7.29) for the spectral density of radiation is called the Rayleigh-Jeans formula. Figure 7.5 shows for comparison the theoretical and experimental curves of spectral density of radiation

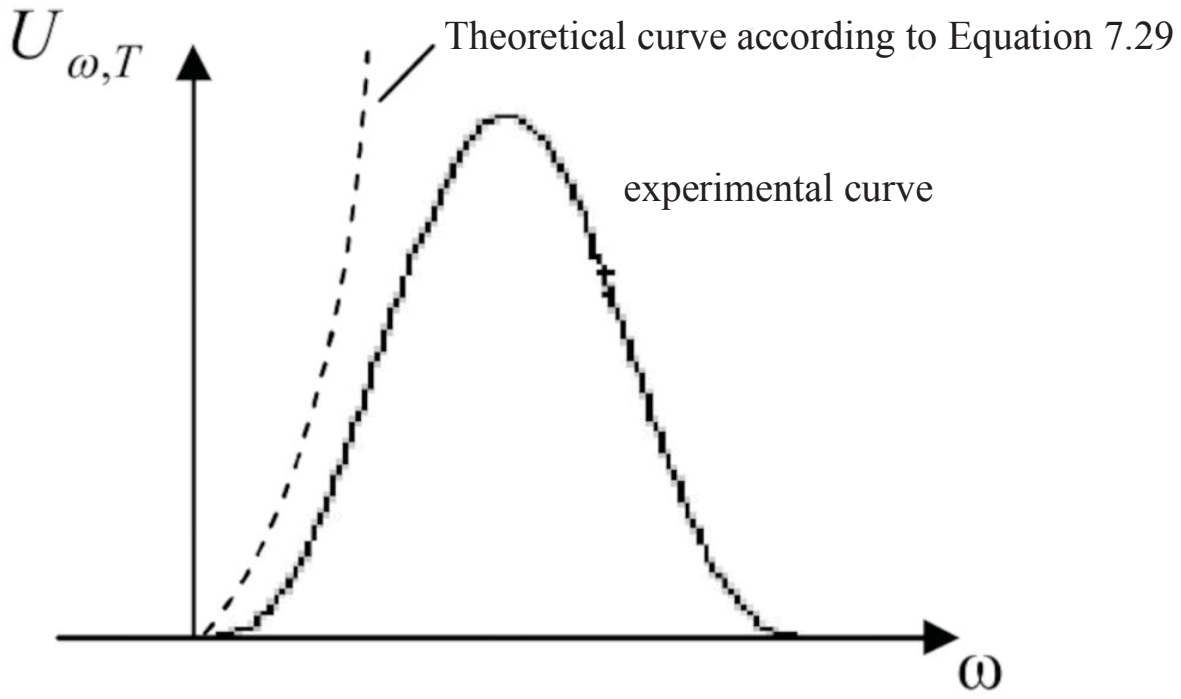


Figure 7.5

Figure (7.5) shows that the Rayleigh-Jeans formula coincides with the experimental curve only in the low frequency region. On the other hand, from (7.29) we see that

$$U_T = \int_0^\infty U_{\omega,T} d\omega \to \infty \qquad (7.30)$$

Thus, the absurd conclusion follows from the Rayleigh-Jeans formula that the energy and mass of a unit volume of thermal radiation are infinite.

The energetic luminosity of a black body, according to its definition, is also infinite, which means that under non-equilibrium conditions the body must instantly give up (emit) all its energy and cool down, respectively, to temperature $T = 0$. All these conclusions from the classical theory contradict the facts following from the experiment, and indicate that the classical theory incorrectly describes thermal radiation. The above situation has been called an ultraviolet catastrophe, as the discrepancy between theory and practice occurred in the high-frequency region. The way out of the crisis was suggested by Planck.

7.2.1.2. Planck's theory of radiation

Planck's theory of radiation is based on his hypothesis, discussed in Sections 6.5.6 and 6.6.3, according to which electromagnetic radiation is emitted and absorbed by discrete portions of energy, photons, for which

$$\varepsilon = \hbar\omega$$

Let, as in the case of classical theory, thermal radiation occurs in some cavity. Taking into account the Planck hypothesis, the radiation in the cavity can be represented by a set of electromagnetic field oscillators with the following natural frequencies

$$\omega_n = n\,\omega, \text{ where } n = 1,2,3,... \qquad (7.31)$$

Thus, the total energy

$$\varepsilon_n = n\hbar\omega \qquad (7.32)$$

An oscillator with energy according to (7.32) is called a **quantum oscillator**.

* In the state of equilibrium, the probability p_n that the oscillator will have energy ε_n that is calculated by the well-known Boltzmann entropy formula

$$p_n = Be^{-\frac{\varepsilon_n}{kT}} \tag{7.33}$$

The constant B is found from the normalization condition, according to which

$$\sum_{n=0}^{\infty} p_n = 1 \tag{7.34}$$

Substituting (7.34) into (7.33) gives

$$B = \frac{1}{\sum_{n=0}^{\infty} e^{-\frac{\varepsilon_n}{kT}}} \tag{7.35}$$

and

$$p_n = \frac{e^{-\frac{n\hbar\omega}{kT}}}{\sum_{n=0}^{\infty} e^{-\frac{n\hbar\omega}{kT}}} \tag{7.36}$$

The average value of the energy of the quantum oscillator is found by the formula

$$\varepsilon_{av} = \sum_{n=0}^{\infty} p_n \varepsilon_n = \frac{\sum_{n=0}^{\infty} n\hbar e^{-\frac{n\hbar\omega}{kT}}}{\sum_{n=0}^{\infty} e^{-\frac{n\hbar\omega}{kT}}} \tag{7.37}$$

Let us denote

$$\frac{\hbar\omega}{kT} = x \tag{7.38}$$

then

$$\varepsilon_{av} = -\hbar\omega \frac{d}{dx} \ln \sum_{n=0}^{\infty} e^{-nx} \tag{7.39}$$

In mathematics, it is proved that the series $\sum_{n=0}^{\infty} e^{-nx}$ converges (see Appendix 2), with ħ

$$\sum_{n=0}^{\infty} e^{-nx} = \frac{1}{1-e^{-x}} \tag{7.40}$$

It is, in this way,

$$\varepsilon_{av} = \frac{\hbar\omega}{e^{\frac{\hbar\omega}{kT}} - 1} \tag{7.41}$$

In the classical limit (see 7.27)

$$\hbar\omega \approx kT$$

In this case, the value of $\hbar\omega$ in (7.41) can be replaced and we can assume that

$$\varepsilon_{av} \approx \frac{kT}{e-1} \approx kT$$

This also follows from the expansion of $e^{\frac{\hbar\omega}{kT}}$ in a series, where we can be limited to the first two terms of the series. Then according to (7.41)

$$\varepsilon_{av} \approx \frac{\hbar\omega}{1 + \frac{\hbar\omega}{kT} - 1} \approx kT.*$$

The above allows us, for the transition from the classical to quantum theory, to replace the value of kT with the value of

$$\frac{\hbar\omega}{e^{\frac{\hbar\omega}{kT}} - 1} \tag{7.42}$$

Thus, to calculate the spectral density, we can use formula (7.29), replacing kT in it with (7.42).

In other words, taking into account (7.29) and (7.41)

$$U_{\omega,T} = \frac{\omega^2}{\pi^2 c^3} \frac{\hbar\omega}{e^{\frac{\hbar\omega}{kT}} - 1}. \qquad (7.43)$$

The relation (7.43) is called the **Planck's law**. Planck's law coincides completely with the experimental curve shown in Figure 7.5. This allows us to state that Planck's hypothesis about the corpuscular nature of electromagnetic radiation can be considered proven. The experimental curve of Figure 7.5 allows calculating the Planck constant $\hbar$ with great accuracy (see Section 6.6.3).

7.2.2. Compton scattering

One of the proofs of the corpuscular nature of electromagnetic radiation was obtained in 1922 by Compton in his experiment on the scattering of X-rays by free electrons. A schematic of the Compton experiment is shown in Figure 7.6. A narrow beam of monochromatic radiation with a wavelength λ is directed from the X-ray source S by the aperture D to the scattering substance B. The scattered radiation is analyzed by spectrograph C, which determines its spectral composition.

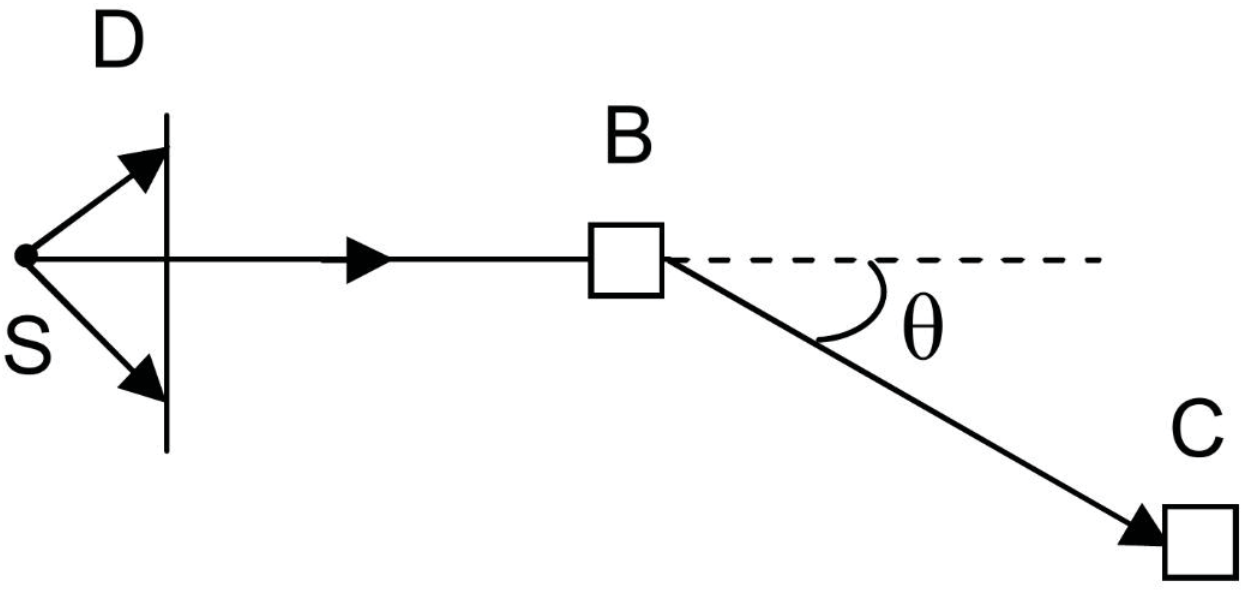


Figure 7.6.

The Compton experiment found that

$$\Delta\lambda = \lambda'(\theta) - \lambda = \lambda_c(1 - \cos\theta), \qquad (7.44)$$

where θ is the scattering angle;

λ is the wavelength of incident radiation;

λ'(θ) is the wavelength of scattered radiation;

λ_c is the Compton wavelength of the electron.

The change in wavelength when scattering electromagnetic waves according to formula (7.44) is called the **Compton effect**. This effect contradicts the classical notion that the wavelength of incident and scattered light is the same. If, however, we attribute momentum to the light wave and apply the law of conservation of momentum, then the relation (7.44) is confirmed.

7.2.3. Photoelectric Effect

The photoelectric effect was discovered by Heinrich Hertz in 1887 and studied by Russian physicist Aleksandr Stoletov in 1887-1889.

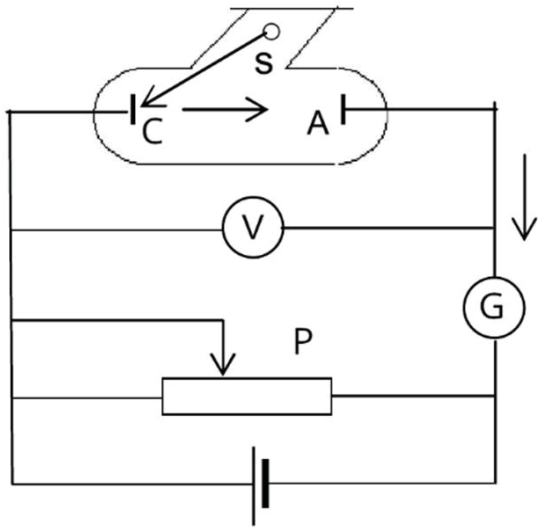


Figure 7.7

Already in the process of these studies, it was discovered that the photoelectric effect does not obey the basic provisions of Maxwell's classical electrodynamics.

The schematic diagram of the photoelectric effect setup is shown in Fig. 7.7. 7.7. Monochromatic light from the source S of frequency ω illuminates the cathode C, which is made of the material under study. This light knocks out the electrons of the upper shells of the cathode atoms (photoelectrons). These electrons under the influence of the anode A create a current I (photocurrent) in the photocell circuit, which is measured with the galvanometer V. By varying the voltage in the circuit with the potentiometer G, take volt-ampere characteristics for different light flux Φ (Fig. 7.8). When the voltage between anode and cathode is $U = 0$, only a small fraction of photoelectrons with low velocities reaches the anode.

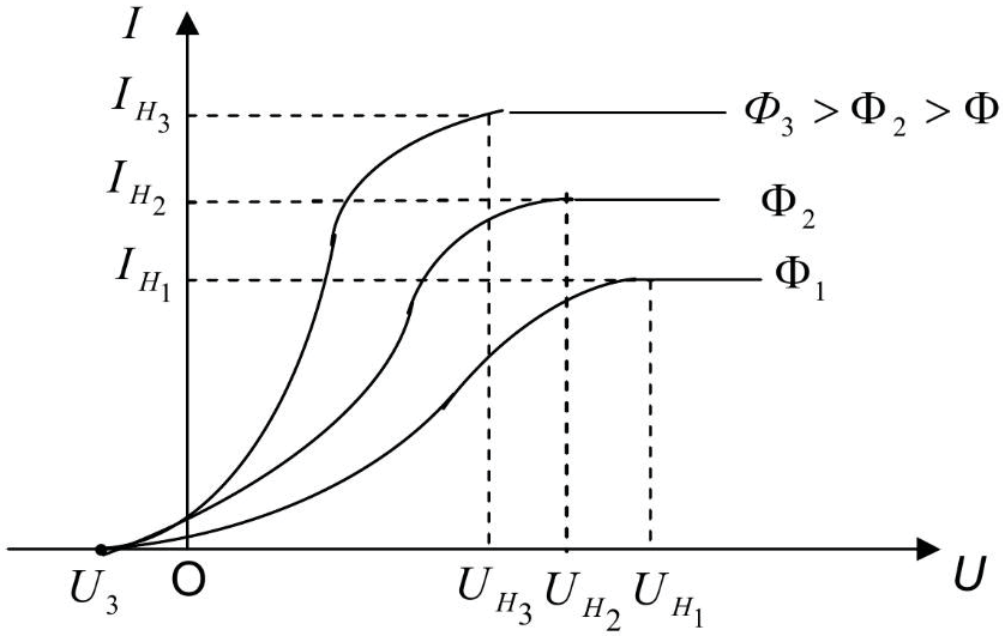


Figure 7.8.

As the voltage increases, the number of photoelectrons reaching the anode increases rapidly, and at a certain value of $U = U_H$, when all the photoelectrons emitted by the cathode per unit time reach the anode,

the photocurrent reaches saturation. Applying a negative potential to the anode, which creates a delaying voltage Ur, reduces the photocurrent to zero. This requires that

$$\frac{\mathrm{m}_0 \upsilon_m^2,}{2} = eU_r \qquad (7.45)$$

where m_0 is the mass of the electron;

υ_m is the maximum velocity of photoelectrons;

e is the charge of the electron.

According to classical electrodynamics, the energy of a photoelectron does not depend on the frequency of light, but is proportional to the light flux (the square of the amplitude of the electric field strength of the light wave). As it increases, according to (7.45), the value of U_r should also increase. However, in fact, as can be seen from the volt-ampere characteristics (Fig. 7.8), the value of U_r does not depend on the value of the light flux Φ, but increases sharply with increasing frequency of monochromatic light. Moreover, it turns out that at frequencies lower than a certain critical frequency defined for each material

$$\omega < \omega_{cr};\ \lambda > \lambda_{cr} \qquad (7.46)$$

the photoelectric effect does not occur at all. In other words, the phenomenon of the photoelectric effect, as well as thermal radiation, confirms the fact of the "ultraviolet catastrophe".

Meanwhile, this discrepancy can be easily explained by Einstein's theory discussed above, which, based on the wave-particle duality, considers light as a flow of photons (see Section 6.6.3). In this case, according to the law of conservation of energy and Einstein's theory

$$\mathcal{E} = \hbar\omega - \mathcal{E}' \qquad (7.47)$$

where $\mathcal{E}$ is the kinetic energy of the photoelectron;

$\mathcal{E}'$ is the ionization energy of the atom, equal to the yield work W.

It follows from relation (7.47) that when

$$\hbar\omega < \mathcal{E}' \tag{7.48}$$

then

$$\mathcal{E} < 0 \tag{7.49}$$

which is impossible. It follows that

$$\omega_{cr} = \frac{\mathcal{E}'}{\hbar} \tag{7.50}$$

This is fully confirmed by experiment.

Since the kinetic energy according to Einstein's relation (7.4.7) depends on frequency and is independent of illumination (at $\omega = 0$ photons are absent and $\mathcal{E} = -\mathcal{E}'$), it is determined only by the yield work, constant for each given substance. The value of U_r in this case does not depend on illumination and increases with increasing frequency.

Thus, the phenomenon of the photoelectric effect also fully confirms the wave-particle duality of light.

7.2.4. Radiation of the atom. Bohr's Postulates

The concept of the smallest indivisible particle, the atom, has been known since ancient times. However, it was discovered only in the 19th century that atoms are the elements that form the simplest chemical substances. At the same time, it was found that the atom, in turn, is a complex particle with its own internal structure. In 1897, the English physicist J.J. Thomson, observing the so-called cathode rays, came to the conclusion that they are a flux of negatively charged particles that are part of the atom, with a very small mass $m_0 \approx 9 \cdot 10^{-31}$ kg. He called these particles electrons. He also created the first model of the atom. According to Thomson's model, the atom consists of very small elementary particles, electrons, which neutralize the positive charge that occupies the entire volume of the atom, within which the electrons float. In 1903 the

physicists Ernest Rutherford and Frederick Soddy discovered that the radioactive radiation of the uranium atom, discovered before that by the French physicist Henri Becquerel, consists of three components, α-rays, which is a flow of fully ionized helium atoms, β-rays, a flow of electrons, and the hard (high-energy) electromagnetic radiation, γ-rays. In 1918 Rutherford investigated the scattering of α-particles on atoms of heavy elements and found the failure of Thomson's model. It turned out that all the positive charge of the atom is concentrated in a minuscule part of its volume and forms a particle, which Rutherford called the **nucleus of the atom**. The size of the nucleus, calculated by Rutherford from the results of the experiment, was only 10^{-15} m. The size of the atom, equal to 10^{-10}, turned out to be a hundred thousand times larger. The entire, virtually empty, volume of the atom, according to Rutherford, is occupied by electrons, which rotate under the action of Coulomb forces in different orbits around the nucleus. The number of electrons is equal to the positive charge of the nucleus. This model of the atom was called the **Rutherford model**.

Meanwhile, Rutherford model of the atom, in terms of its radiation, came into conflict with experience. According to this model, the electrons rotating around the nucleus, moving with acceleration, would, according to Maxwell's electromagnetic theory (see Section 4.2.3), continuously emit electromagnetic waves, forming a solid spectrum of radiation. Experience, however, showed that the emission spectrum of atoms is linear. In addition, the electrons of the atoms would have to inevitably fall on the nucleus due to their continuous emission, giving the atoms extreme instability, which was also in contradiction with experience.

Direct experiments performed in 1913 by the German physicists James Franck and Gustav Hertz also showed that the energy of the electrons in the atom cannot be any and takes only a certain number of discrete values, which also contradicted classical electrodynamics.

Linear spectra of hydrogen-like atoms were discovered as early as 1890 (**hydrogen-like or one-electron atoms** are ionized atoms with one electron, see section 4.1.2.2). The spectral series of these atoms are described by the following formula obtained experimentally, namely

$$\frac{1}{\lambda_{mn}} = RZ^2\left(\frac{1}{n^2} - \frac{1}{m^2}\right), \tag{7.51}$$

where λ_{mn} is the wavelength observed in the spectrum;

$m > n$ are positive integers, not equal to zero, defining the order of the two allowed energy levels;

R is the Rydberg constant;

Z is the charge of the atomic nucleus.

With this in mind, Bohr supplemented Rutherford's classical model of the atom in 1913 with the following two postulates, which relied on the wave-particle duality introduced by Planck and Einstein, and brought the theory of the atom in line with experiments:

1) there are states of the atom with corresponding discrete energy values $\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_n, ...$, and being in these states, the atom neither radiates nor absorbs energy;

2) the atom absorbs and emits intermittently only when it transitions from one state to another; the frequency ω_{mn} of the emitted or absorbed radiation at transitions is determined by Planck's formula

$$\omega_{mn} = \frac{\mathcal{E}_m - \mathcal{E}_n}{\hbar}, \tag{7.52}$$

The quantization rule introduced by Bohr, which singles out the stationary state of the atom, also follows from Planck's formula. According to Planck's formula, the momentum of the electron

$$P = nk\hbar.$$

Since

$$k = \frac{2\pi}{\lambda}, \text{ and } P = m\upsilon,$$

then

$$m_0\upsilon = n\frac{2\pi}{\lambda}\hbar,$$

from which

$$m_0\upsilon = n\frac{2\pi}{\upsilon T}\hbar = n\frac{2\pi\omega}{2\pi\upsilon}\hbar = \frac{n\hbar}{r},$$

Finally

$$m_0\upsilon r = n\hbar \tag{7.53}$$

The value $m\upsilon r$ is the angular momentum, hence, the relation (7.53) is the condition of momentum quantization. From (7.53) we obtain that the Bohr radius of the orbit is

$$r = \frac{n\hbar}{m_0\upsilon}. \tag{7.54}$$

The velocity υ can be determined according to Rutherford's model, namely, based on the fact that the centripetal force is equal to the Coulomb force, i.e.

$$\frac{m_0\upsilon^2}{r} = \frac{Ze^2}{4\pi\varepsilon_0 r^2}. \tag{7.55}$$

From here

$$\upsilon^2 = \frac{Ze^2}{m_0 4\pi\varepsilon_0 r^2}. \tag{7.56}$$

and

$$r_n = \frac{4\pi\varepsilon_0 n^2\hbar^2}{Zm_0e^2}.$$

On the other hand, the total energy of an electron in the field of a stationary nucleus

$$\mathcal{E} = \frac{m_0 v^2}{2} - \frac{Ze^2}{4\pi\mathcal{E}_0 r}. \tag{7.57}$$

where the second term is the potential energy U

$$U = \int F_K dr = -\frac{Ze^2}{4\pi\mathcal{E}_0 r}.$$

Substitution of v^2 and r_n into (7.55) give

$$\mathcal{E}_n = -\frac{1}{n^2} \cdot \frac{m_0 e^4 Z^2}{32\pi^2 \mathcal{E}_0^2 \hbar^2}. \tag{7.58}$$

It follows from (7.58) that Bohr's postulates for the spectral series coincide with formula (7.51) obtained experimentally, provided we assume that the Rydberg constant

$$R = \frac{m_0 e^4}{64\pi^3 \mathcal{E}_0^2 c\hbar^3} \approx 1.097 \cdot 10^7 m^{-1}.$$

This quantity coincides with the experimental value of the Rydberg constant with a fairly high accuracy. This means that Bohr's theory is confirmed experimentally.

The Bohr theory is, nevertheless, semi-classical and fully deterministic, because it leaves the laws of classical electrodynamics and mechanics unchanged, only introducing, in accordance with the Planck-Einstein hypothesis, quantization of electron energy and momentum and the idea of discrete states in which the electron does not emit and does not absorb. Based on the hypotheses of Planck and Einstein, Louis de Broglie in 1923 suggested that Bohr's postulates are a consequence of wave properties exhibited by particles (see Section 6.6.4) and derive from a more general physical theory, which was later called **quantum theory**.

7.3. From classical physics to quantum theory

The facts considered, which were already discovered at the beginning of the last century, showed that classical physics was not applicable to a number of phenomena of interaction between radiation and matter and to the processes occurring at the microlevels. Investigations of these facts eventually led to the creation of a new, quantum physics. Despite the discovered deviations in the behavior of photons and electrons from classical theory, scientists who studied these deviations and stood on purely deterministic positions, including Planck, Einstein, Louis de Broglie, tried to harmonize them with classical theory by creating a number of new postulates supplementing classical theory.

Moreover, starting from classical deterministic considerations (see section 6.4), using the postulate of wave-particle duality (equations 6.29 and 7.1), the relations called subsequently as uncertainty relations (see section 7.3.1) were obtained. The resulting uncertainties from the point of view of de Broglie and his school, however, do not indicate any violation of the principle of determinism, but are purely statistical in nature.

In other words, they are, according to de Broglie, the consequence of statistical processes that inevitably arise in observations of microparticles due to the irreducible interactions of these particles with the photons of the light beam. De Broglie, in this connection, expressed the hope as early as the middle of the 20th century that uncertainties serve only as a measure of the degree of ignorance of microprocesses occurring at deeper structural microlevels. At the same time, he believed that with the penetration of science to these levels, these uncertainties would disappear.

The situation changed considerably in 1927, when the German physicist Heisenberg obtained the same uncertainty relations not based

on classical concepts, but solely on the postulated indeterministic properties that objectively belong to microparticles.

Heisenberg showed that uncertainty ratios characterize a single particle rather than collectives of particles, as de Broglie believed, and therefore are not statistical, but objectively natural, arising from internal properties of the particle. Indeterminism, according to Heisenberg, is inherent in microparticles as well as all physical bodies from nature.

In other words, indeterminism is the fundamental law of Nature. It is, however, not observed in macroprocesses due to its exceptionally small size.

Today, as will be shown later, it is already quite obvious that the dispute between de Broglie and Heisenberg, which lasted until the 1970s, was unequivocally resolved in favor of Heisenberg's concept.

7.3.1. Uncertainty Relations

Uncertainty relations are a fundamental law of Nature, underlying not only quantum theory, but all physics. They are at the same time a direct consequence of the wave-particle duality of all material formations without exception.

According to Schrödinger (see Section 6.6.4.1), a free particle can be assigned a wave function $\psi(x)$, the square of which is determined by the probability of finding the particle at a point with coordinate x. With this

$$\psi(x+\lambda)=\psi(x) \tag{7.59}$$

All functions of the following type are known to have this property

$$\psi(x)\sim e^{jkx} \tag{7.60}$$

Indeed, let us consider the function

$$e^{jk(x+\lambda)}.$$

It is obvious that

$$e^{jk(x+\lambda)} = e^{jkx}e^{jk\frac{2\pi}{\lambda}} = e^{jkx}e^{j2\pi} = e^{jkx}$$

Thus, the state of a free particle with a certain momentum P and energy $\mathcal{E}$ is described by the de Broglie wave function

$$\psi(x) = Ce^{j\frac{P}{\hbar}kx} \tag{7.61}$$

The square of the modulus of the function $\psi(x)$, as seen from (7.61), is independent of x, since it is determined by the square of the amplitude of this function.

$$|\psi(x)|^2 = |C|^2 \left|e^{j\frac{P}{\hbar}kx}\right| = |C|^2$$

In other words, a particle with a strictly defined momentum (energy) is not localized. This means that it can be detected at any point in space with probability $|\psi(x)|^2$. Such a particle, like a monochromatic wave with a strictly defined wavelength (frequency), is an idealization and does not exist in actual reality. The real particle corresponds not to a monochromatic de Broglie wave, but to a packets of de Broglie waves. Such an approach does not contradict the initial positions, since the solution of the Schrödinger wave equation (see Section 6.6.4.1) corresponds, as is known from mathematics (see Appendix 2), not only to the function $\psi(x)$ defined by equation (7.61), but also to any superposition of such functions.

Mathematically the specified superposition is defined by the Fourier integral (see Appendix 2), which in the limit describes a signal strictly localized in space. The following relations apply (see formula 6.3) for this signal

$$\left.\begin{aligned}\Delta k \cdot \Delta x &\approx a\\ \Delta\omega \cdot \Delta t &\approx b\end{aligned}\right\} \tag{7.62}$$

From these relations, in particular, it follows that

$$\left.\begin{array}{c} \Delta k \to \infty \; and \; \Delta\omega \to \infty \\ \Delta \mathrm{x} \to 0 \; and \; \Delta \mathrm{t} \to 0 \end{array}\right\} \tag{7.63}$$

The relations (7.63) mean a strict localization of the considered wave packet in space and time.

From a physical point of view, a particle to which a wave packet with wavelengths in the Δk interval and frequencies in the $\Delta\omega$ interval corresponds is localized. This is a consequence of the fact that near some fixed value of $x = x_0$ or $t = t_0$, where the phases of the monochromatic waves coincide with each other, all the amplitudes add up and give a resulting amplitude not equal to zero. Away from the values of $x = x_0$ ($t = t_0$) at

$$|x - x_0| \gg \lambda; \; |t - t_0| \gg T$$

the phases of the added waves have a huge difference, the amplitudes mutually cancel each other out so that the resulting amplitude is on average equal to zero.

Combining further the equations (7.62 and 7.1), we conclude that

$$\left.\begin{array}{c} \Delta \mathrm{P} \cdot \Delta \mathrm{x} \approx \hbar \\ \Delta \mathcal{E} \cdot \Delta \mathrm{t} \approx \hbar \end{array}\right\} \tag{7.64}$$

where ΔP, $\Delta \mathcal{E}$, Δx and Δt are possible intervals of change of the specified quantities.

The relations (7.64) were, as mentioned above, called **uncertainty relations**. If ΔP, $\Delta \mathcal{E}$, Δx and Δt are understood as their standard deviations from the mean values (i.e. the variance of these waves), then the uncertainty relations (7.64) take the form

$$\left.\begin{array}{c} \Delta \mathrm{P} \cdot \Delta \mathrm{x} \geq \dfrac{\hbar}{2} \\ \Delta \mathcal{E} \cdot \Delta \mathrm{t} \geq \dfrac{\hbar}{2} \end{array}\right\} \tag{7.65}$$

It follows from the uncertainty relations that the more precisely one of the quantities in an inequality is determined, the less certain is the value of the other. No experiment can change this situation, since the uncertainty is not related to the level of perfection of measurement technique, but to objective, inherent in nature, properties of the particles under study themselves. It is very important to understand that these properties are universal and belong to all real-world systems without exception. However, in the macrocosm dominated by macro-processes, which are characterized by sufficiently large space-time intervals, uncertainties remain unnoticed due to their smallness.

In this regard, we are confronted with a completely new reality that cannot be observed in the macrocosm around us. It was this circumstance that led to the need for a new physics describing a world completely different from the one in which we live.

Let us consider the main characteristics of this physical reality.

7.3.1.1. Space-time symmetry

It follows from the uncertainty relations that the postulates about the homogeneity and isotropy of space-time, considered inviolable in classical physics, derived from notions of space-time symmetry, are erroneous and contradict the reality. In fact, physical laws are in direct correlation with the dimension of space-time intervals, which becomes more and more evident as we go deeper into matter.

The fundamental, eternal and invariable from the point of view of classical physics laws of conservation of energy, momentum and angular momentum, following from the classical notions of homogeneity and isotropy of space-time, are an idealization of actual reality, and in fact have no place. For small space-time intervals, as follows from the uncertainty relations, there are relatively large uncertainties of momentum and energy, which determine the nature of microprocesses. For example, intra-atomic electromagnetic processes have a duration of

the order of 10^{-16} s. This means that the energy uncertainty for them is of the order of

$$\Delta\mathcal{E} \approx \frac{\hbar}{\Delta t} = \frac{1.05 \cdot 10^{-34}}{10^{-16}} = 1.05 \cdot 10^{-18}\, J \approx 6.6\, eV$$

For intra-atomic processes, this is a fairly significant amount of energy, sufficient, for example, for the self-excitation of the atom. The uncertainty energy gives rise to the tunnel effect (see Section 6.6.4.1), which underlies many microprocesses, including the beta decay of radioactive nuclei, autoionization of the atom in a strong electric field, thermonuclear reactions, autoelectronic emission, phenomena in the contact layer at the boundary of two semiconductors, and a number of other effects. All these processes, from the point of view of classical physics, violate the laws of conservation of energy and momentum. These laws are observed in them only on average, i.e., when the observation time is long enough.

The tunnel effect, in particular, occurs when the time required for the particle to jump over the barrier is sufficient for the resulting energy uncertainty to exceed the height of the barrier. On average, this does not violate the law of conservation, because the energy of the particle jumping over the barrier, according to the same uncertainty relations, decreases over time back to the initial value.

7.3.1.2. Stability of the atom

Another example of the effect of uncertainty relations in the microcosm is the problem of the stability of the atom. Let us consider this problem with the example of the hydrogen atom. Let the electron rotate around a nucleus (proton) along a circular orbit with radius r with constant velocity v in accordance with Rutherford's classical model. The Coulomb force of attraction of the electron to the nucleus, equal to

$$\frac{e^2}{4\pi\varepsilon_0 r^2},$$

is balanced, according to the second law of dynamics, by the centripetal force

$$\frac{m_0 \upsilon^2}{r}$$

From here

$$\mathrm{r} = \frac{e^2}{m_0 \upsilon^2 4\pi\varepsilon_0}. \qquad (7.66)$$

Thus, if the speed of the electron is large enough, approaching the speed of light in a vacuum, the radius of its orbit tends to the radius of the nucleus.

Another reason why the electron must fall on the nucleus is that, rotating around the nucleus with centripetal acceleration υ 2 / r, it must, according to classical electrodynamics, continuously radiate and, therefore, move not along a circular but a spiral orbit, constantly approaching the nucleus. Substituting the value of velocity from the first uncertainty relation into the above equation for the electron orbital radius, we obtain

$$\mathrm{r} \geq \frac{16\pi\varepsilon_0 \hbar^2}{m_0 e^2} \approx 2 \cdot 10^{-10}\ m.$$

Thus, the minimum orbital radius of the electron cannot be less than $2 \cdot 10^{-10}$ m. The smallest possible orbital radius is provided in this case by the second uncertainty relation, according to which during the motion of the electron along the above orbit one revolution is performed on average for a time $\Delta t \approx 4.5 \cdot 10^{-17}$ c. This time interval corresponds to an energy uncertainty of about 7.5 eV, which balances the loss of energy

that would occur in 1 revolution. As a result, the atom does not radiate, and the electron remains in orbit.

The uncertainty relations thus allowed us, for the first time, to estimate the dimensions of the atom, which coincided with the dimensions obtained with high accuracy from experiments unrelated to the motion of the electron around the nucleus. They also allowed us to explain why the atom, contrary to classical laws, is stable, and - to understand why the electrons, being in a stationary orbit, do not move toward the nucleus. By the way, already from the uncertainty relations even before the discovery of the Pauli exclusion principle, which will be discussed below, it follows that they prevent the convergence of particles of matter so that these particles do not merge. In other words, for the first time it was possible to explain that due to the action of uncertainty relations the discreteness of structures at all levels and the extent of real systems in the Universe is ensured.

7.4. Fundamentals of Quantum Mechanics

Depending on the field of research, quantum physics is conventionally divided into three large sections - quantum mechanics, quantum field theory, and quantum statistics. In addition, it includes applied sections that reference related fields and use its laws for practical purposes. Applied quantum physics, in turn, includes quantum electrodynamics, quantum electronics, quantum optics, quantum chemistry, quantum energy, etc.

Since all sections are based on the concepts and quantities established in quantum mechanics, let us first consider its basic provisions.

Quantum, or otherwise, **wave mechanics** is a physical theory that is based on a wave-particle dualism. It studies the laws of motion of microparticles, including atoms, atomic nuclei, elementary particles, and

the systems they comprise (e.g., crystals, semiconductors, superconductors, etc.). Quantum mechanics also studies the relationship between quantities that characterize these objects and quantities measured directly in the experiment.

Laws of quantum mechanics, as it was mentioned above, apply in principle not only to microparticles and their systems, but to all objects of material world without exception. It follows that classical mechanics and its laws can be regarded as a particular, limiting case, for which quantum effects appear so insignificantly that they can usually be neglected.

On the other hand, since the properties of macroscopic bodies are the result of averaging of interactions occurring at microlevels, most macroscopic phenomena find an explanation only within the framework of quantum mechanics.

Quantum mechanics allowed, for example, to explain the structure and properties of macromolecules, the properties of solids, including metals, dielectrics and semiconductors, the phenomenon of ferromagnetism, superfluidity, superconductivity, a number of optical phenomena, chemical processes, and so on.

The laws of quantum mechanics underlie many major technical achievements, including nuclear and thermonuclear energy, the theory and practice of radiation, materials science, as well as the purposeful search and creation of new materials with unique properties, etc.

Classical mechanics is characterized by describing the motion of bodies by specifying their position in space at each moment of time, which corresponds to their movement along well-defined trajectories. However, as we already know, such a description does not correspond to reality and is applicable only when considering bodies of sufficiently large mass. In quantum mechanics, the motion of bodies is not described by the equation of motion, but by the Schrödinger wave equation, the

solutions of which are reduced to finding the wave function of the free and bound states of particles. As follows from these solutions, the states of particles at each moment of time are characterized not by their coordinates and time, but by a set of so-called quantum numbers. The consequence of this, in particular, is the absence of an orbital bound electron in the atom. In other words, the notion of moving toward it is inapplicable. In addition, the electron density of the electron is, by one law or another, scattered all over the atom. If we apply the classical concepts, we can talk not about the movement of the electron inside the atom (rotation around the nucleus), but about the pulsation over time of its density, or, more precisely, about the pulsation of its probability of being at a given point at a given time.

Quantum mechanics is divided into relativistic and non-relativistic mechanics. Since relativistic quantum mechanics is closely related to field theory, here we limit ourselves to considering only non-relativistic quantum mechanics.

Non-relativistic quantum mechanics is a semi-classical theory. It allows a conditional description of the state of particles using classical concepts - motion, forces, orbits, as well as Rutherford's classical model of the atom, etc. In this case, the problems of electron motion inside atoms under the action of central forces are of particular importance. This approach is called the **correspondence principle**.

7.4.1. Quantum theory of the hydrogen-like atom

One of the most remarkable successes of quantum mechanics was the theory of the hydrogen-like atom. This theory is based on Schrödinger wave equations and is in brilliant consistency with experiment. In particular, it made it possible to solve the problem of electron motion in a hydrogen-like atom. Let us first consider this problem for the classical hydrogen atom. The nucleus of a hydrogen atom consists of one

positively charged particle. The value of the charge of this particle is assumed to be 1, and the particle itself is called a proton.

Rutherford's experiments on the scattering of α-particles, which have already been mentioned above, made it possible to calculate with great accuracy the ratios between the masses of the proton m_p and the electron m_e. It turned out that almost all the mass of the hydrogen atom is concentrated in its nucleus, that is, the proton, so that m_p / m_e = 1836. Taking this circumstance into account, we will further assume that the proton is fixed at the center of the atom and connected with a conditionally stationary reference frame. In this frame of reference the proton can be seen as a source of a spherically symmetric field, and the electron as a test charge introduced into this field. In this case, we consider the proton and electron to be point-like, i.e. we assume that their interaction occurs according to Coulomb's law. The potential energy of the electron $U_p(r)$ in the Coulomb field is known to be

$$\mathrm{U_p(r)} = -\frac{Ze^2}{4\pi\varepsilon_0 r}. \qquad (7.67)$$

where Z is the charge of the atomic nucleus;

r is the distance between the nucleus and the electron;

e is the charge of the electron.

Graphically, the relation (7.67) has the form shown in Fig. 7.9., where $U_p(r)$ is the potential energy of the electron, and $\mathcal{E}$ its total energy.

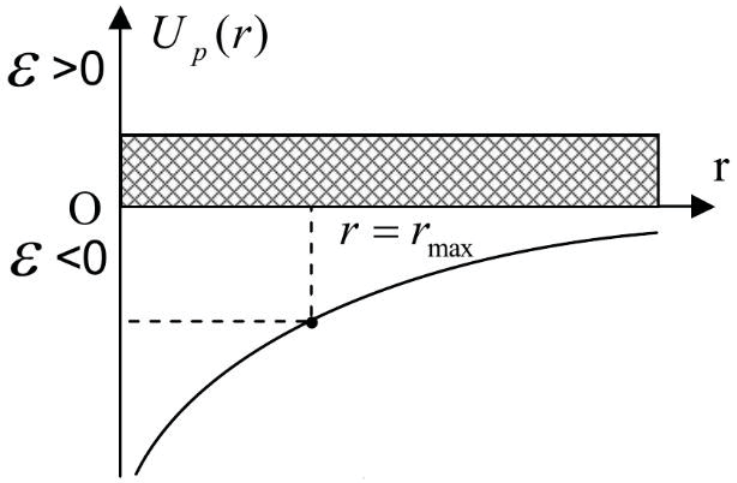


Figure 7.9

Thus, within the framework of the classical theory, the motion of the electron bound to the nucleus is finite (see Section 4.1.1.2). The region $r \leq r_{max}$ on the graph corresponds to the condition of bound (finite) motion of the electron, in which its kinetic energy W_k is less than or equal to the binding energy $|U_p|$, so that its total energy

$$\mathcal{E} = W_k - |U_p| \leq 0.$$

This means that during the finite motion of the electron, the radius of its orbit r cannot exceed r_{max}. Indeed, when $r > r_{max}$ $\mathcal{E} > 0$, and $W_k > |U_p|$. Such an electron is detached from the atom and becomes free, and the atom is ionized.

As can be seen from the graph, the energy spectrum of the electron in the classical theory at any values of energy is continuous (in Figure 7.9 this state is depicted by shading).

The electron in the potential well of the hydrogen atom behaves quite differently in the framework of the theory of quantum mechanics. Its motion in this case is described by the wave function ψ, which is a solution of the stationary Schrödinger wave equation. In Section 6.6.4.1, this problem was considered in the Cartesian coordinate system.

However, the spherical coordinate system is more convenient. In this system, the solution of the Schrödinger equation ψ (r, φ, θ) can be represented as the product of three functions ψ_1 (r), ψ_2' (φ), and ψ_2'' (θ), which are called, respectively, **radial**, **azimuthal**, and **orbital**.

This solution is discussed in more detail in Problem 3, Section 7.7. In this problem it is shown that in addition to the principal quantum number *n* (see Section 6.6.4.1), which characterizes the radial function, the azimuthal function is described by the magnetic quantum number m_l, and the orbital function by the orbital quantum number *l*. The specified quantum numbers take integer values.

The quantum number m_l is similar to the wave number *k* in a plane wave. The azimuthal function characterizes the rotational part of the electron's motion in the atom, such as the rotation of the electron around the Oz axis.

Rotational motion is similarly described in classical mechanics.

The vector of angular momentum $\vec{L}$ in classical mechanics, as we know,

$$\vec{L} = [\vec{r} \times \vec{P}]. \tag{7.68}$$

The component of this vector on the Oz axis, L_Z, describes the rotational motion around this axis.

Since for a plane wave

$$k = \frac{P}{\hbar}, \tag{7.69}$$

then for the magnetic quantum number by analogy

$$m_l = \frac{L_Z}{\hbar}. \tag{7.70}$$

From relation (7.70), taking into account the perpendicularity of the velocity vector, and hence the momentum, to the position vector r ($\vec{r} \perp \vec{P}$), and also from relation (7.68) it follows that

$$\Delta L_i = \Delta x_i \cdot \Delta P_i, \text{ where } (i = x, y, z) \qquad (7.71)$$

From (7.71) and the uncertainty relations (7.65) we obtain that

$$\Delta L_X \cdot \Delta L_Y \cdot \Delta L_Z \sim \hbar_3 \qquad (7.72)$$

Thus, in quantum mechanics it is impossible to set the orbital angular momentum by exact values of its three projections.

If one of the projections is exact, as follows from (7.72), the other projections are indeterminate.

It can be shown further (without proof) that, according to Correlation 4.1 (see Section 7.7), the absolute value of the orbital momentum L_l can be given by the exact value of any of its projections. In doing so, it turns out that

$$L_l = \sqrt{l(l+1)} \cdot \hbar \qquad (7.73)$$

where l, like other quantum numbers, form a discrete series related to the principal quantum number by the relation

$$l = 0, 1, 2, \ldots, (n-1) \qquad (7.74)$$

It also turned out that the orbital and magnetic quantum numbers are related, namely

$$m_l = 0, \pm 1, \pm 2, \ldots, \pm l \qquad (7.75)$$

Thus, it follows from the solutions of the Schrödinger equation that the state of the electron in a hydrogen-like atom is characterized by the three quantum physical quantities mentioned above, namely:

- energy, which is determined by the quantum number n;

- orbital momentum momentum L_l, which is determined by the orbital quantum number l;

- by the projection of this momentum L_{lz} on the Oz axis, which is determined by the quantum number m_l.

In addition to the orbital angular momentum $\vec{L}_l$, the electron also has an intrinsic angular momentum $\vec{L}_S$. It is conventionally assumed that this momentum is caused by the rotation of the electron around the axis, not related to its orbital motion, and is called the spin momentum. In principle, a mechanical model of the electron, explaining the emergence of the orbital and eigenmomentum of its momentum can be the formation from a solid shell and liquid contents. As such an electron rotates, its internal content lags behind due to friction.

As in the case of the orbital momentum, the projection of the L_{SZ} spin momentum can by analogy be given by the quantum relation

$$L_{SZ} = m_S \hbar \tag{7.76}$$

where m_S **is the quantum magnetic spin number**.

The spin number is sometimes referred to simply as **spin**. The spin of an electron inside an atom can take only two values, which are considered antiparallel.

$$m_S = \pm 1 / 2. \tag{7.77}$$

The quantum numbers n, l, m_l, and m_S form a system that fully characterizes the state of the electron inside the hydrogen atom and hydrogen-like atoms.

7.4.1.1. Analysis of the solution of the Schrödinger equation

Let's calculate the number of possible states i of an electron in an atom, defined by different combinations of quantum numbers. Since for each value of the principal quantum number n the orbital number l takes values from $j = 0$ to $j = (n - 1)$, then

$$i = \sum_{j=0}^{j=n-1} l_j. \tag{7.78}$$

On the other hand, for each value of l_j the orbital number m_l takes $(2j + 1)$ values. ml = 0; ± 1; ± 2; and the spin number mS takes two values $m_S = \pm ½$.

It is, in this way,

$$lj = 2\,(2j + 1). \tag{7.79}$$

Substituting l_j from (7.79) into (7.78) gives

$$i = \sum_{j=0}^{j=n-1} 2(2j + 1). \tag{7.80}$$

from which

$$i = 2\,[1 + 3 + 5 + \ldots + (2n - 1)]. \tag{7.81}$$

The expression in square brackets of the last ratio is an arithmetic progression with the common difference d = 2, so

$$i = 2\frac{1 + 2n - 1}{2} \cdot n = 2n^2. \tag{7.82}$$

The energy level $\mathcal{E}_n$, which corresponds to a single quantum state ($i = 1$) is called **non-degenerate** (ground state). Levels for which $i \geq 2$ are called **degenerate**. The number of states corresponding to a given value of n is called the **degree of degeneracy**. Since each quantum number n and its energy level $\mathcal{E}_n$ corresponds to its conditional electron orbital radius, states with a given quantum number n form electron layers. The first electron layer closest to the nucleus is denoted by the letter K, the second by L, the third by M, etc. In each layer there are different states with different values of the orbital number l.

For the K-layer, $n = 1$, $l = n$ - 1, and, respectively,

$$l = 0;\ m_l = 0;\ m_S = 2;\ i = 2n^2 = 2.$$

For the L-layer, $n = 2$ and, respectively,

$$l = 0;\ m_l = 0, \pm 1;\ m_S = 2;\ i = 2n^2 = 8.$$

For the M-layer, $n = 3$ and, respectively,

$$l = 0;\ m_l = 0, \pm 1, \pm 2;\ m_S = 2;\ i = 2n^2 = 18.$$

Let us denote, as it is common in quantum mechanics, the states with a given number l by s, p, d, f, ...-states. It is not difficult to see that the wave functions corresponding to the s-state with $l = 0$ are spherically symmetric. This applies equally to states with any $n = 1, 2, ...$ and $l = 0$. These spherically symmetric functions are respectively denoted ψ_{1S}, ψ_{2S}, ψ_{1S}, ...

Let us consider, for example, the function ψ_{1S} ($n = 1$, $l = 0$) . Since $|\psi_{1S}|^2$ is the value of the probability of detecting an electron at distance r from the nucleus, the value $4\pi r^2|\psi_{1S}|^2$ determines the probability of detecting an electron in a spherical layer of unit thickness located at distance r. These probabilities are shown in Fig. 7.10a. Fig. 7.10b shows a spherical layer of radius r. The figure shows that the electrons inside the atom are distributed in a spherical region surrounding the atomic nucleus, with the maximum electron density occurring at the Bohr radius r_b. The shaded area on either side of the value of the Bohr radius (Fig. 7.10b) corresponds to the maximum probability of electron detection.

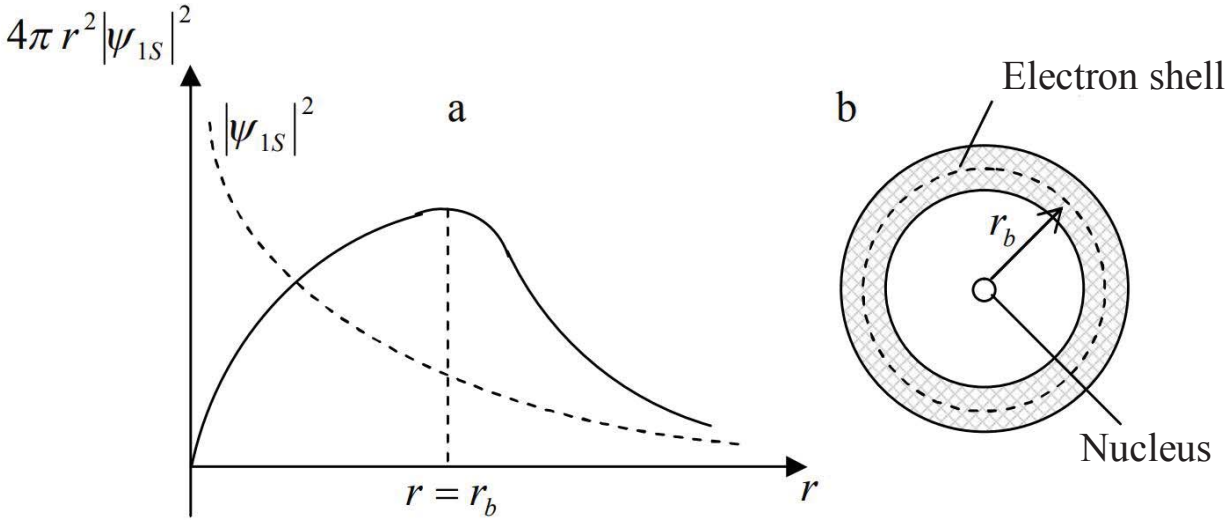


Figure 7.10.

From Figure 7.11, where the curves of the functions ψ_{jS}, we see that they intersect the r axis $(j - 1)$ times.

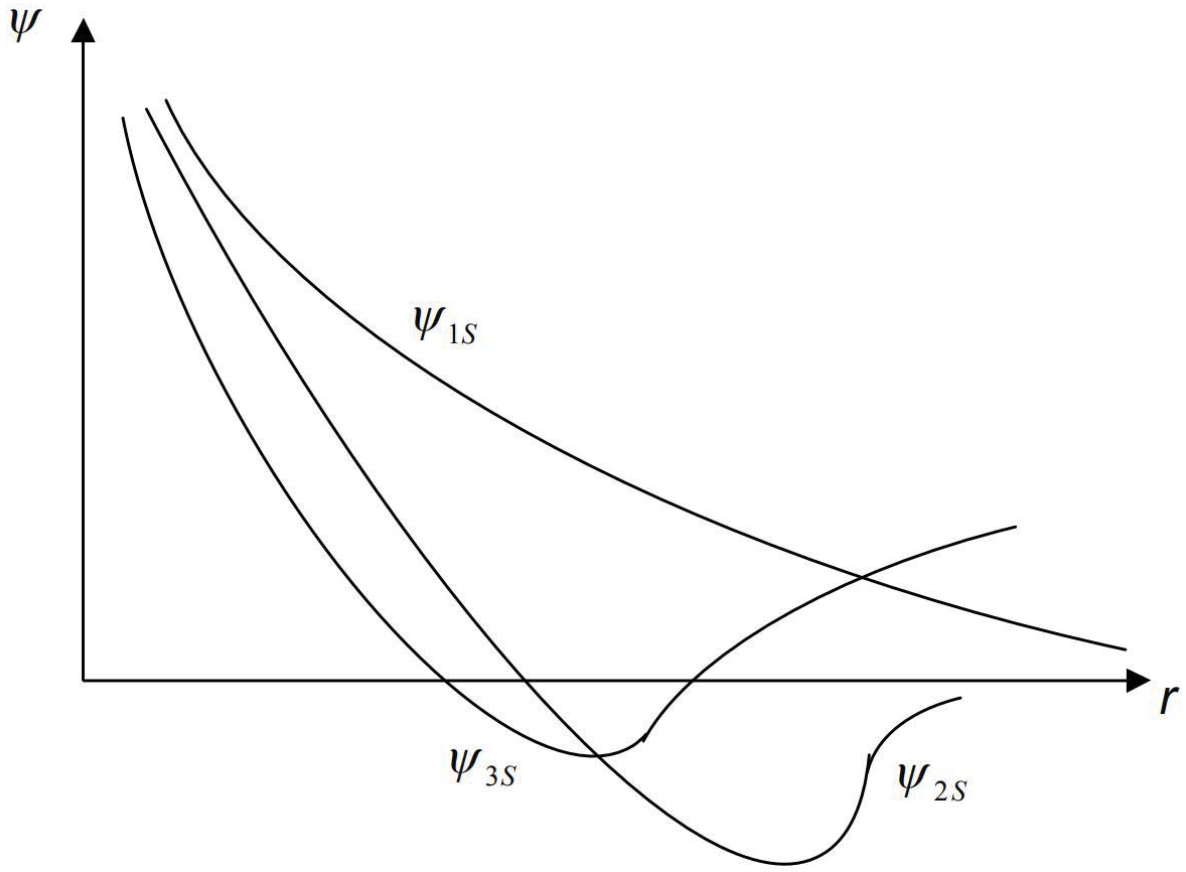


Figure 7.11.

The intersection points correspond to spherical surfaces with corresponding radius r_j. The probability of detecting an electron at these points is zero. These surfaces are called nodal surfaces. For the 1s-state $j = 0$; for the 2s-state j = 1, 3s – j = 2, and so on.

Figure 7.12a shows, respectively, the curves of the dependences of $|\psi_{2S}|^2$ and $4\pi r^2|\psi_{2S}|^2$ on r and the electron density distribution.

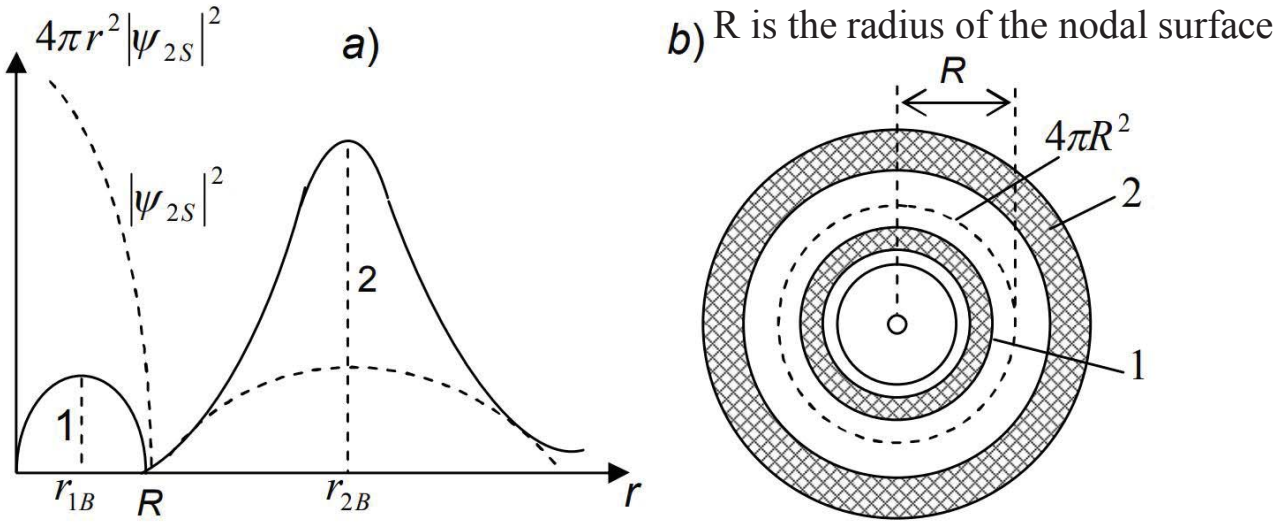


Figure 7.12.

As can be seen in Figure 7.12 b, electron 2s can be in two spherical layers, 1 and 2, separated by the nodal surface $4\pi R^2$ (dashed line). The 1st layer of the 2s-state overlaps the layer of the 1s-state (unshaded area near layer 1). However, the probability of detecting a 2s-electron in layer 2 is much higher than in layer 1 (see Figure 7.12a). Thus, the size of the atom in the 2s state is larger than in the 1s state.

The 1s state corresponds to the stationary state of the atom, and the other states are 2s, 3s, ... are corresponding to excited states. They are all spherically symmetric and differ from each other in the degree of excitation and the number of nodal surfaces. Since the 1s-state is stationary, i.e. $\Delta t \to \infty$, then, according to the uncertainty relation, $\Delta \mathcal{E} \to 0$. It follows that a stationary state corresponds to a well-defined energy level and the classical deterministic behavior of the atom, as well as the

observance in it of the laws of conservation of energy and momentum. The energy level $\mathcal{E}_1$ is infinitely subtle. On the contrary, the energy levels of excited states $\mathcal{E}_2$, $\mathcal{E}_3$, ... are blurred. For them, during the time $\Delta\tau_i$ the electron stays at the excitation level, $\mathcal{E}_i \approx \hbar / \Delta\tau_i$.

In the $p, d, f...$ - states of the electron, for which $l \neq 0$, its detection probability is determined not only by the distance of the corresponding layer to the nucleus, but also by the angles θ and φ. Solutions of the Schrödinger wave equation lose their spherical symmetry in this case. At the same time, the electron density distribution and the shape of the electron cloud change. Thus, for example, analysis of the solution of the equation for $l = 1$ (states) shows that these solutions are also symmetric about the corresponding coordinate axes, but in the form of dumbbells. They are also characterized by the number $(n - 2)$ of nodal surfaces. Figure 7.13a shows dependence plots of the $2p$ and $3p$ states, as well as electron density distribution diagrams of the $2p$ state (7.13b), for which $(l = 1)$ and m = 0, ± 1.

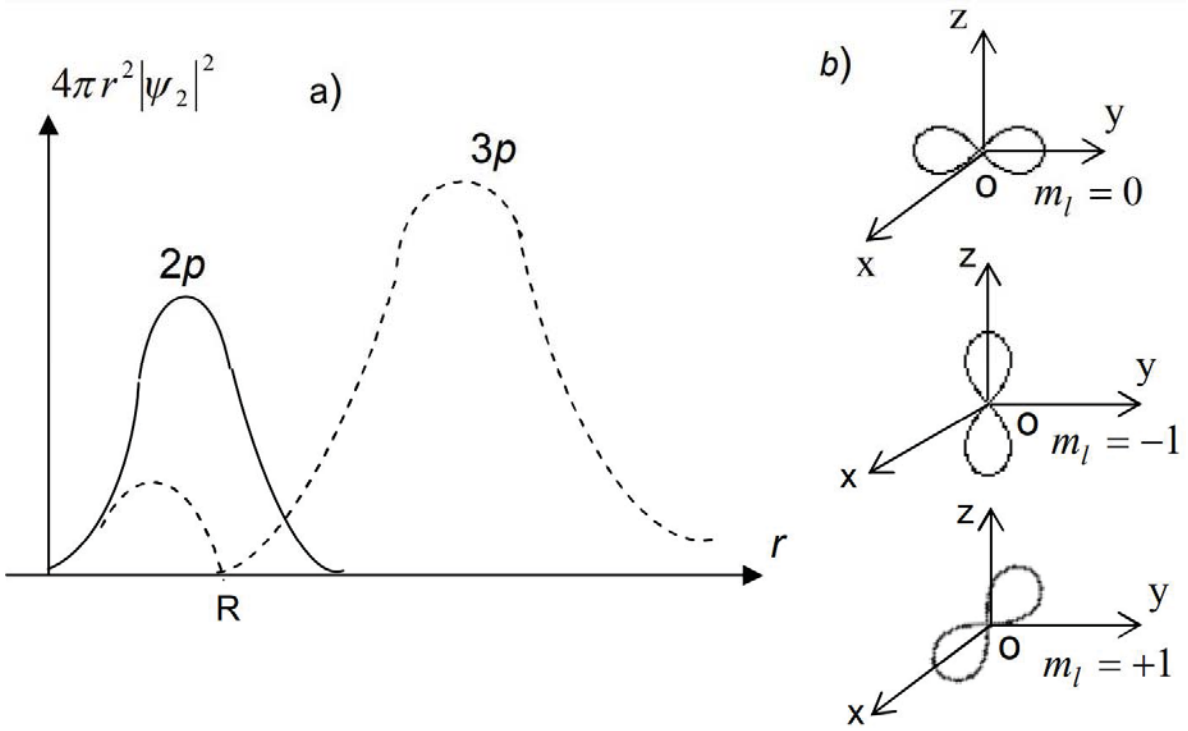


Figure 7.13.

It is interesting to note that the total distribution function describing the probability of detecting an electron in any of the $2p$ states is spherically symmetric. In the nitrogen atom, for example, $Z = 7$. There are only 7 electrons in the electron shells of this atom, including 2 of them placed in the K-layer, in the $1s^2$ state ($l = 0$; $m_l = 0$; $m_S = 2$) . Five electrons are placed in the next M-layer. Two electrons are in the $2s^2$ state ($l = 0$; $m_l = 0$; $m_S = 2$), and three are in the $2p^3$ state (ml = 0; ± 1).

Symbolically, the states of the nitrogen atom are therefore written in the form

$$1s^2\ 2s^2\ 2p^3 \qquad (7.83)$$

The $1s^2$ and $2s^2$ states are spherically symmetric by definition, and the $2p^3$ state is spherically symmetric because it contains all 3 $2p$ states, so the nitrogen atom as a whole is spherically symmetric. The wave function in the d-state ($l = 2$) has (n - 3) nodal surfaces and is similar in configuration to the p-state. It has a higher degree of symmetry, and the electron density distribution also has the form of a four-lobed rosette. For example, in the d-state, where there are no nodal surfaces yet, and $m_l = 0$; ± 1; ± 2, the electron density has the following distributions (Fig. 7.14).

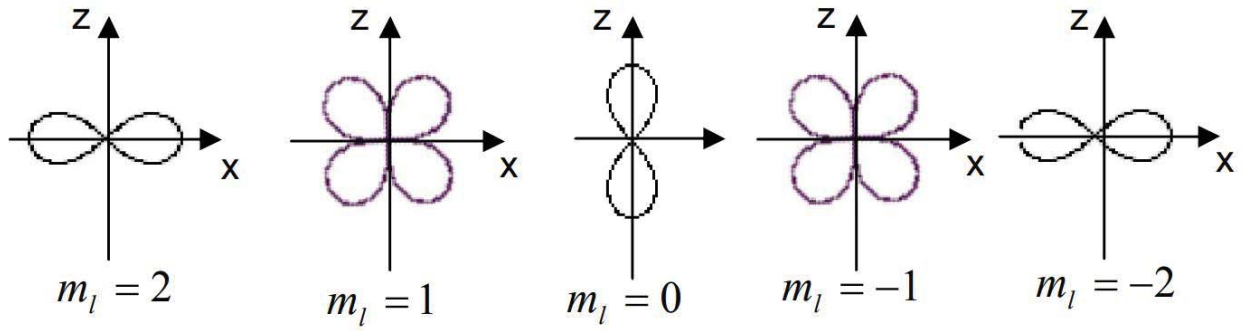


Figure 7.14

The quantum state of the total distribution function at the position of all five d-states is the same as in the case of p-states, it is spherically symmetric. For example, a manganese atom (Z = 25) contains a spherically symmetric inner K-shell (2 electrons), a spherically

symmetric L-shell (8 electrons), all 5 spherically symmetric *d*-states (5 electrons) and 2 spherically symmetric *s*-states (2 electrons - outer shell). It follows that the wave function of the manganese atom is generally spherically symmetric.

When an atom has fewer than three *p*-states or fewer than five *d*-states, and each *p*- or *d*-state has more than two electrons, it is not spherically symmetric, but is stretched in one direction or another.

The electron density distribution in *f*, *g*, and other states can be represented similarly.

7.4.2. Systems with a large number of particles

The exact solution of the Schrödinger equation for a system with a large number of particles, such as a multielectron atom, is associated with serious mathematical difficulties. Therefore, the solution of such a problem is simplified by applying a number of physically justified assumptions with a sufficiently high approximation. These assumptions follow from an analysis of the properties of microparticles based on general considerations and reliable observations.

7.4.2.1. Principle of Identity

The principle of identity is a purely quantum effect arising from quantum-mechanical consideration of properties of microparticles. In classical mechanics, as we know, all bodies, including microparticles, are considered as corpuscular entities, which are strictly localized in space at any given moment of time in the chosen frame of reference, characterized by the exact value of the coordinate, and their motion is described by well-defined paths. Therefore, absolutely identical particles (for example, two electrons) are fundamentally different in their location in space-time. In quantum mechanics, the state of particles is described by a wave function, which can only be used to determine the probability of their location. For example, let particle 1 move relative to some reference

frame and collide with particle 2 stationary relative to this frame. From the point of view of classical mechanics, if the velocity and direction of motion of particle 1 are known, then it is possible to determine how the trajectories of particles 1 and 2 will change after their collision and to distinguish particle 1 from particle 2 by these trajectories. From the point of view of quantum mechanics, particles have no trajectory and changing their location cannot be experimentally detected. This statement is the principle of identity of quantum mechanics.

Let us consider a system of two particles. Let the first particle be described by the wave function $\psi_1 (q_1)$ and the second particle by $\psi_2 (q_2)$. Arguments of functions q_1 and q_2 are generalized coordinates characterizing quantum numbers n, l, m_l, m_S. Let $\psi_1 (q_1, q_2)$ be the wave function describing a two-particle system in which the first particle is in state 1 and the second is in state 2. It follows from the principle of identity that

$$|\psi_1 (q_1, q_2)|^2 = |\psi_1 (q_2, q_1)|^2$$

From here

$$\psi_1 (q_1, q_2) = \pm \psi_1 (q_2, q_1) \qquad (7.84)$$

A wave function that satisfies the first condition, i.e., the equation (7.84) with plus sign, is called an **even symmetry** function. A wave function that satisfies the second condition, i.e., the equation (7.84) with a minus sign, is called an **odd symmetry** function.

The parity property of symmetry is a characteristic feature of this kind of particles. If the wave function of a particle has odd symmetry, then any 2 such particles cannot be in the same states. On the contrary, if the wave function of a particle has even symmetry, then any number of them can be in the same state. The number of non-interacting particles that are in the same state with the same set of quantum numbers under given conditions is called the **occupation number**. Thus, the occupation numbers for particles with even symmetry can only take values 0 and 1.

The rule that forbids particles with odd symmetry to be in the same state is called the **Pauli exclusion principle**.

7.4.2.2. Relation between parity and spin of particles

Earlier it was mentioned that in quantum mechanics the electron is attributed with its own angular momentum. This angular momentum only provisionally characterizes the motion of the particle. In fact, it has a purely quantum character and is not related to its displacement. The concept of electron eigenmomentum was introduced into physics based on the analysis of its spectroscopic data in 1925 by American scientists George Eugene Uhlenbeck and Samuel Abraham Goudsmit. They interpreted this moment as the result of the rotation of the electron around the axis like a spinning top. Hence the name of this new quantum characteristic, **spin**, which means to rotate, to spin.

Spin was introduced into the mathematical apparatus of quantum mechanics in 1927 by the Austrian physicist Wolfgang Ernst Pauli, with the so-called Pauli equation. This equation is a generalization of the Schrödinger equation for the electron, which takes into account its spin. In 1928 Paul Dirac derived the Pauli equation from the general relativistic electron equation as its particular non-relativistic approximation. Subsequently, it was found that all particles, both simple and complex, have spin. Spin is detected by observing the spectra of particles by their so-called **fine structure**. The latter is a consequence of the interaction of the magnetic field of the eigenmomentum (spin) of particles with the magnetic field of their orbital moment. For the first time the fine structure, i.e. **splitting of energy levels**, was observed by the Dutch physicist Pieter Zeeman back in 1896 (Zeeman effect) when studying the glow of sodium vapor in a magnetic field. The Zeeman effect consists in the fact that under the action of a magnetic field, three lines (triplets) are observed in the radiation spectra of atoms instead of one spectral line in the direction perpendicular to the magnetic field strength. In more complex cases, doublets (two lines) and multiplets (more than three lines)

are also observed. The Zeeman effect is explained by the interaction of the magnetic moment arising in a quantum system under the action of an external magnetic field with the magnetic moment of the atom. The number of splitting lines is determined by the degree of energy level degeneracy. The fine structure of particles is considered conventionally as a result of the interaction of the magnetic field of the orbital momentum of the Coulomb rotation of a particle (electron) around the nucleus with the magnetic field of its own momentum (spin). This interaction leads to the formation of an energy addition that splits the energy levels. The intrinsic momentum of the particles is similar to the orbital momentum and obeys the same relations (see 7.71, 7.74). This means that one of the projections of the eigenmomentum is also quantized, i.e.

$$L_{CZ} = m_S \hbar \tag{7.85}$$

where L_{CZ} is the projection of the eigenmomentum on the Oz axis;

m_S is the spin quantum number.

In addition, the absolute value of the eigenmomentum

$$L_C = \hbar^2 m_S (m_S + 1) \tag{7.86}$$

The number of values the eigenmomentum projection can take, as in the case of the orbital momentum,

$$k_S = (2m_S + 1) \tag{7.87}$$

where k is the multiplicity of spin momentum degeneracy.

It turned out that the value of k_S takes on values of arbitrary integers and zero

$$k_S = 0; \pm 1; \pm 2 \tag{7.88}$$

From relation (7.87) it follows, therefore, that for odd values of k the spin m_S is zero or an integer, and for even values it is a half integer ± 1 / 2, ± 3 / 2, ± 5 / 2, etc. In the particular case of the electron

$$m_S = \pm \frac{1}{2} \tag{7.89}$$

Particles are only conventionally regarded as spinning tops. A more clear and close to reality representation of spin is given by information related to particle symmetry (Fig. 7.15)

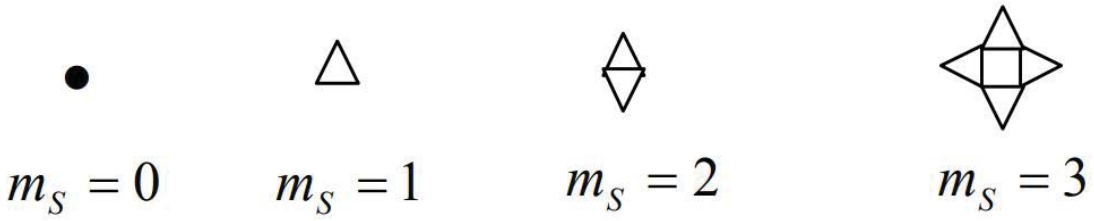


Figure 7.15.

A particle with spin equal to zero resembles a point with all-round (absolute) symmetry in this sense, because it looks the same from all sides. A particle with spin 1 can be compared to a one-sided arrow, which looks different from different sides. However, if you rotate it 360º, it will look the same. A particle with spin 2 can be compared with a symmetrical double-sided arrow. It takes on its former appearance if you turn it 180º. Particles with higher integer spin take on the same appearance when rotated to even smaller angles, for example, if $m_S = 3$, to an angle of 90º, etc. However, there are also such particles which do not possess the above symmetry, since when they are completely rotated by 360º they do not take their former form. This happens only if they are rotated by 360º twice, which is connected with a half-integer value of spin of these particles.

Thus, all known particles in the Universe can be divided into two groups:

- particles whose wave function is odd; they have half-integer spin; they are fermions that form atomic matter; they include the following particles and their antiparticles - electrons, nucleons, quarks, neutrinos. They also include complex particles formed by an odd number of fermions; fermions are subject to the Pauli exclusion principle;

- particles whose wave function is even; they have integer or zero spin; these are bosons that carry interactions and do not obey the Pauli exclusion principle; they include photons, gluons, vector bosons, gravitons, mesons, and complex particles composed of an even number of fermions. The connection between parity and spin was first established in 1940 by Pauli.

7.4.3. Periodic Trends of the Structure of the Atom

The periodic trends was discovered by the Russian chemist Dmitri Mendeleev. It turned out that if the simplest one-atom chemical substances (elements) are arranged in order of increasing their atomic masses, then after a certain number of elements that form an interval or period, there is a repetition of their chemical and physical properties. If the elements with similar properties are arranged in ascending order of their atomic masses one below the other in vertical columns, all elements form eight such columns. These columns were called groups and designated by Roman numerals from I to VIII. The first group included hydrogen and the single-valent metals of the alkaline series (H, Li, Na, K, Rb, Cs, Fr). The second group included the divalent metals of the alkaline-earth series (Be, Mg, Ca, Sr, Ba, Ra), etc. As we move from one group to another in each period, the valence increases and the chemical activity decreases, and the metallic properties are gradually replaced by non-metallic ones. Elements with clearly expressed non-metallic properties are located in groups IV - VIII. From group IV to VIII the valence decreases from 4 to 0, the non-metallic properties and chemical activity increase and reach their maximum in the elements of group VIII, the so-called halogens. Group VIII is formed by inert gases with zero valence and chemical activity.

As a result, all the elements are arranged so that they form a rectangular table called the Periodic Table. Horizontal rows of the Periodic table create periods. Each period accommodates elements with

different, gradually changing properties. Each period except the first begins with a single-valent, highly reactive metal and ends with an inert gas. The first period includes only two elements, H and He. The second and third periods include 8 elements each from Li to Ne and from Na to Ar, respectively. The fourth and fifth periods include 18 elements each. The fourth period begins with K and ends with Kr, and the fifth with Rb and Xe. The sixth period includes 32 elements from Cs to Rn, and the seventh, incomplete period, includes only 19 elements. It begins with Fr and ends with the last relatively stable 105 element of the table, Db (Dubnium).

Let us consider how the periodic trends relate to the charge of the nucleus and the orderly arrangement of the electrons within each atom. First, let us point out that all elements are arranged in the Periodic table not as their atomic mass increases, as Mendeleev believed, but as their nucleus charge increases, i.e., the number of protons in the nucleus and the number of electrons in the atom corresponding to them. In addition to protons, neutrons are also found in the nucleus, which do not affect its charge. The mass of the neutron is almost the same as that of the proton. The number of neutrons in light and medium nuclei is approximately equal to the number of protons, and in heavy nuclei there are considerably more of them. The mass of atoms increases with the charge of the nucleus, but the periodic trends and the dependence on mass are observed only approximately and have many exceptions. The periodic law is completely conditioned by the Pauli exclusion principle.

Let's call electrons with the same values of the principal quantum number n equivalent. The degree of degeneration of atoms with the same numbers n is known to be

$$2\,(2l + 1)$$

The number of equivalent electrons with a given value of l forms, as mentioned above, electron shells, which are denoted respectively by

s(l = 0); p(l = 1); d(l = 2); f(l = 3), etc. The number of electrons located on the electron shell is indicated on the top right with an Arabic numeral. For example,

s means that 1 electron is located in the s-shell,

s^2 means that there are 2 electrons,

p^3 means that there are 3 electrons on the p-shell, etc.

Each shell can accommodate a limited number of electrons, namely:

on the s shell - 2 electrons;

on the p shell - 6 electrons;

on the d shell - 10 electrons;

on the f shell - 14 electrons, etc.

Shells containing the specified number of electrons are called completely filled shells. Thus, s^2, p^6, d^{10}, f^{14}, etc. - are the designations of completely filled shells. It is obvious that there can be no shells designated, for example, s^3, p^8, or f^{16}, etc.

Equivalent electrons are located in the corresponding electron layers, which are denoted by K(n = 1); L(n = 2); M(n = 3); N(n = 4), etc.

Each layer can contain a maximum number of electrons equal, as mentioned above, to $2n^2$. Layers containing the maximum possible number of electrons are called **completely filled layers**. The value of the principal quantum number is usually indicated before the shell designation (see Table 7.1).

Table 7.1.

Shells		Layers			
Designation	Number of electrons of a completely filled shell	K	L	M	N
		n =1; l =0	n =1;2, l = 0;1	n = 1;2;3, l = 0;1;2	n=1;2;3;4, l=0;1;2;3
s	2	1s -$1s^2$	1s-$1s^2$; 2s-$2s^2$	1s-$1s^2$; 2s-$2s^2$; 3s-$3s^2$.	1s-$1s^2$; 2s-$2s^2$ 3s-$3s^2$; 4s-$4s^2$
p	6	-	2p-$2p^6$	2p-$2p^6$; 3p-$3p^6$	2p-$2p^6$; 3p-$3p^6$; 4p- $4p^6$
d	10	-	-	3d-$3d^{10}$	3d-$3d^{10}$; 4d-$4d^{10}$
f	14	-	-	-	4f-$4f^{14}$
…	…	-	-	-	-
Number of electrons of completely filled layers		2	8	18	32

The K-layer is closest to the nucleus. It is followed by L, M, N, ...layers in order. The bonding strength of an electron in an atom decreases with the distance of the layer from the nucleus, i.e., with an increase in the quantum number *n*.

The shells within each given layer move away from the nucleus as l increases, so the bonding strength of the electron with the nucleus within the layer decreases with increasing distance of the shell

For example, the electrons in the M-layer are weaker bound to the nucleus than in the K-layer, and the electrons in the d-shell of the M-layer are weaker bound to the nucleus than the electrons in the s and p shells of that layer. The last shells of the last layer are called **outer shells**. The electrons of unfilled outer shells are weakerly bound to the nucleus than the electrons of filled shells, and the smaller the filling of the last shell, the weaker this bond. Weakly bound electrons of the outer shells easily

separate from their atoms, fill the interatomic space and form the so-called electron gas.

Outer shells containing more than 3-4 electrons are stable. Their electrons are being bound near the nucleus. Such shells easily attach to themselves the electrons that are missing to fill them completely. The filling of energy levels by electrons occurs so that the levels with the lowest possible energy are filled first, then the higher levels, and so on. As a result, all energy levels in the ground state of the system will be filled, from the lowest level to the highest possible level. Higher levels are left vacant. If an electron moves from a filled level to a free (vacant) level, this is regarded as the birth of a particle-hole pair. If an electron returns to the region of filled levels, filling a hole, this is considered recombination, i.e., the disappearance of the pair or annihilation.

Since annihilation sharply increases the electron's bonding to the filled level region, there is a release of bonding energy (mass defect). Energy is emitted by the field quanta of a given particle.

An atom whose electron shells are unfilled or partially filled forms a **positively charged ion**. As the shells fill, the ion charge decreases. Each ion corresponds to a different filling order designation.

An example of filling electron shells for the atoms of nitrogen and phosphorus is shown in Table 7.2 on the next page.

Sometimes the filling of the shells is intermittent, which consists in the fact that there are processes associated with disturbances in the distribution of energy. For example, after filling the 3p state of the M-layer in argon, the K-layer fills not the 3d state of the same layer, but the 4s state of the next layer, which is energetically more advantageous. Electrons with a lower l but higher n are more strongly bonded than electrons with a higher l but lower n (e.g., electrons of the 4s state are more strongly bonded than those of the 3d state). This leads, among other things, to groups of elements, **lanthanides** and **actinides**. Most of the

properties of the atom are determined by the structure and characteristics of the outer shell electrons. The binding energy of these electrons is relatively low and ranges from a few eV to several tens of eV. For comparison, the binding energy of internal electrons is 10^2 - 10^4 eV. Electrons of outer shells participate in emission processes in the optical and thermal range of frequencies (ultraviolet, visible and infrared radiation), in chemical reactions, creation of intermolecular bonds. It is these that determine the electrical and magnetic properties of elements and substances. The magnetic moments of atoms with completely filled shells are usually zero, while the magnetic moments of atoms with partially filled outer shells determine the para- and diamagnetic properties of matter.

Table 7.2.

Example of filling the atomic shells

Filling the shells of the nitrogen ion							
N^{+7}	N^{+6}	N^{+5}	N^{+4}	N^{+3}	N^{+2}	N^{+}	N
-	1s	$1s^2$	$1s^22s$	$1s^22s^2$	$1s^22s^21p$	$1s^22s^22p^2$	$1s^22s^22p^3$

Filling the shells of the phosphorus ion									
P^{15+}	P^{14+}	P^{13+}	P^{12+}	P^{11+}	P^{10+}	P^{9+}	P^{8+}	P^{7+}	P^{6+}
-	1s	$1s^2$	$1s^22s$	$1s^22s^2$	$1s^22s^21p$	$1s^22s^22p^2$	$1s^22s^22p^3$	$1s^22s^22p^4$	$1s^22s^22p^5$

Continued					
P^{5+}	P^{4+}	P^{3+}	P^{2+}	P^{+}	P
$1s^2 2s^2 2p^6$	$1s^2 2s^2 2p^6 3s$	$1s^2 2s^2 2p^6 3s^2$	$1s^2 2s^2 2p^6 3s^2 3p$	$1s^2 2s^2 2p^6 3s^2 3p^2$	$1s^2 2s^2 2p^6 3s^2 3p^3$

The dependence of the properties of atoms on the binding energy of the electrons of their outer shells leads to the fact that they change dramatically during their ionization, as well as during the formation of chemical compounds in which the electrons of the outer shells create ionic or covalent bonds. For example, the properties of the water molecule H_2O have nothing in common with the properties of hydrogen and oxygen atoms. On the contrary, the electrons of the inner shells of atoms have almost no effect on their properties.

7.4.4. Structure of energy levels

When considering the periodic trends and constructing the Periodic table, it was assumed that each electron in a multi-electron atom behaves independently of the other electrons, interacting only with their total average electrostatic field and the nucleus. The sets of states for a given electron were determined by the values of the principal quantum number and the corresponding mechanical momenta, orbital and spin. This approach did not take into account the spin-orbit interaction, the interaction of electrons with the magnetic field of the nucleus, etc. With this approach, the radiation spectra of atoms were represented as a set of spectral lines, which reflected only the discrete energy levels of the atom and the corresponding transitions of electrons from excited to ground states. In reality, the influence of unaccounted factors leads to a change in the specified approximation of the spectral pattern. These changes manifest themselves as a splitting of the lines of the main spectrum, which is called a **fine** and in some cases **hyperfine structure**.

7.4.4.1. Selection rules

The thin and hyperfine structure of the spectrum of the atom is thus a consequence of a number of factors not taken into account in the approximate spectral scheme. Let us first consider the fine structure caused by the spin-orbit interaction. For this purpose, let us introduce the

vector of the total momentum $\overrightarrow{M}$. It can obviously be found from the equality

$$\vec{M} = \vec{L}_l + \vec{L}_S \tag{7.90}$$

where $\vec{L}_l$ is the total orbital momentum of the electrons of the atom;

$\vec{L}_S$ is the total spin angular momentum of these electrons.

The projection of the total momentum to some direction Oz is determined, as usual, by the quantization condition

$$M_{Zj} = m_{ljs}\,\hbar \tag{7.91}$$

where m_{ljs} is the magnetic quantum number of the projection of the total momentum;

The value of the total momentum in view of (7.91) is determined from the relation

$$M = \hbar\sqrt{l_{jS}(l_{jS}+1)} \tag{7.92}$$

where l_{js} is the orbital quantum number of the total momentum.

M is the absolute value of the total mechanical moment

Obviously, for each electron

$$\left.\begin{aligned} l_{jS} &= l \pm \frac{1}{2}, l = 0, 1, 2, \ldots (n-1) \\ m_{ljS} &= m_i + m_S = m_i \pm \frac{1}{2} \end{aligned}\right\} \tag{7.93}$$

Since only positive values of orbital quantum numbers correspond to energy levels, we obtain from relation (7.93) that at

$$l = 0, l_{jS} = \frac{1}{2};\ \ l = 1, l_{jS} = \frac{1}{2}, \frac{3}{2};\ \ l = 2, l_{jS} = \frac{3}{2}, \frac{5}{2} \tag{7.94}$$

Equation (7.94) shows that the same non-zero value of the orbital quantum number corresponds to two values of l_{jS}.

This means that each level, defined by a non-zero orbital number, splits into three sublevels. This determines the fine structure characterized by multiplet

$$t = (2 L_S + 1). \qquad (7.95)$$

It should also be borne in mind that each total mechanical moment corresponds to its own magnetic moment.

The fine structure for transitions from level $l = 1$ to level $l = 0$ is shown in Figure 7.16.

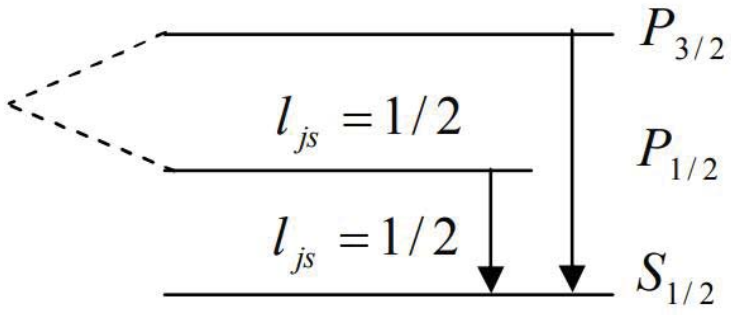


Figure 7.16

Figure 7.16 shows that transitions from level $l = 1$ to the non-dissociated level $S_{1/2}$ ($l = 0$) correspond to a doublet of two close states $P_{3/2}$ ($l_{js} = 3 / 2$); $P_{1/2}$ ($l_{js} = 1 / 2$).

For example, for the sodium atom $n = 3$ and $l = 0, 1, 2$, the inner layers K and L are completely filled, and the layer M contains 1 electron. The excitation levels of this electron are vacant. Since $n = 3$, $l = 2$ and $l_{jS} = 2 \pm \frac{1}{2}$.

The energy structure of the sodium atom is written as

$$1s^2\ 2s^2\ 2p^6\ 3d^1$$

The D level splits into two sublevels, $D_{5/2}$ and $D_{3/2}$, and forms a duplet of yellow lines with wavelengths of 589 and 589.6 Å.

The effect of the emergence of a fine structure in the absence of an external magnetic field is called the **internal Zeeman effect**.

In addition to the fine structure, the spectra of atoms also exhibit a **hyperfine structure** caused by the interaction of the magnetic moment of the electron with the weak magnetic field of the atomic nucleus.

In light and medium atoms, where the spin-orbit interaction is small compared to the electrostatic interaction, the orbital and spin moments of individual electrons add up to each other. In this case, multielectron atoms can be considered as one-electron atoms with one total orbital $\vec{L}_l$ and one spin momentum $\vec{L}_S$, for which

$$\left.\begin{aligned} \vec{L}_l &= \sum_i \vec{L}_{li} \\ \vec{L}_S &= \sum_i \vec{L}_{Si} \end{aligned}\right\} \qquad (7.96)$$

If the inner shells are completely filled and their moments are compensated, then we are limited to considering only the outer shell electrons (valence electrons). If the number of valence electrons is even, their total spin quantum number is an integer; if the number is odd, it is half an integer.

In the central electric (Coulomb) field, electrons conditionally rotate around the nucleus and their total moments $\vec{L}_l$ and $\vec{L}_S$, while remaining constant in magnitude, continuously change their orientation (direction). The vector of the resulting total momentum $\vec{M}$ is preserved in both magnitude and direction. This pattern corresponds to the rotation of the vectors $\vec{L}_l$ and $\vec{L}_S$ with angular velocity ω around the vector $\vec{M}$.

If the spin-orbit interaction is large compared to the electrostatic interaction of electrons with each other, for example, in heavy atoms, then the relation (7.96) is not satisfied. In this case, the following takes place for each i-th electron

$$\left.\begin{aligned}\vec{M}_i &= \vec{L}_{li} + \vec{L}_S \\ \vec{M} &= \sum_i \vec{M}_i\end{aligned}\right\} \tag{7.97}$$

The state of an atom is determined not only by the total momentum $\vec{M}$, but also by each momentum $\vec{M}_i$ and its corresponding quantum number.

When an electron of a complex atom transitions from one state to another, it can emit or absorb a quantum of electromagnetic energy.

If there are transitions of external electrons, there is radiation in the optical spectrum. Transitions of internal electrons emit radiation in the X-ray spectrum.

Let us consider, for simplicity, single-photon processes. This approach is acceptable because multiphoton processes are unlikely. If $\vec{M}_B$ is the total angular momentum of the electron (electrons) before the radiation, $\vec{M}_E$ is the total angular momentum after the radiation, and $\vec{S'}_P$ is the vector of the photon's own angular momentum, then according to the law of conservation of angular momentum (the photon has no orbital momentum),

$$\vec{M}_B = \vec{M}_E + \vec{S'}_P \tag{7.98}$$

The spin of the photon is known to be 1. If we also assume that at the initial moment

$$\left|\vec{M}_B\right| = 0, \text{ then } \left|\vec{M}_E\right| = 1 \tag{7.99}$$

$$\left|\vec{M}_E\right| = 0, \text{ then } \left|\vec{M}_B\right| = 1 \tag{7.100}$$

In other words, if in the initial or final state one of the momenta $\vec{M}_B$ or $\vec{M}_E$ is zero, then the possible transitions in the emission and absorption of a photon are only those at which

$$\Delta M = |M_B - M_E| \tag{7.101}$$

Transitions from $M_B = 0$ to $M_E = 0$ are forbidden.

In the first case, when the vectors $\left|\vec{M}_B\right| \neq 0$ and $\left|\vec{M}_E\right| \neq 0$ and are directed along the same line, they must have the opposite orientation. Otherwise, $\vec{S'}_P \leq 0$, which contradicts the original assumption. In the second case, when vectors $\vec{M}_B$ and $\vec{M}_E$ are directed at an angle, the vector equality (7.98) can be realized only if $\left|\vec{M}_B\right| = \left|\vec{M}_E\right| = 1$ (see Figure 7.17). In this case, the value of $\Delta\vec{M} = 0$.

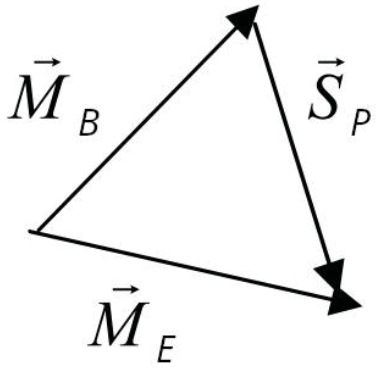


Figure 7.17

If $\left|\vec{M}_B\right| = -\left|\vec{M}_E\right|$ (opposite orientation), then $\Delta\vec{M} = \pm 1$. All other transitions are forbidden.

Thus, summarizing the above, we obtain for the absolute values of momenta

$$\left.\begin{array}{c} \Delta\vec{M} = \pm 1; 0 \; at \; \vec{M}_B \neq 0 \; or \; \vec{M}_E \neq 0 \\ \Delta\vec{M} = \pm 1 \; at \; \vec{M}_B = 0 \; or \; \vec{M}_E = 0 \\ \Delta m_j = 1; 0 \end{array}\right\} \qquad (7.102)$$

The equations (7.102) are called the **selection rule**. The emission of light is possible either as a result of a change in orbital momentum $\vec{L}_l$, or as a result of rotation $\vec{L}_S$, or both, for example, by the interaction of a photon with the magnetic moment of an atom. However, when emitting in the optical spectrum, this interaction is much smaller than the interaction of the photon with the charge of the electron, causing a change

in $\vec{L}_l$. In this case we can assume that $\vec{L}_S = 0$. Therefore, the selection rules can be rewritten as follows

$$\left.\begin{array}{l} \Delta L_l = 0; \pm 1 \quad at\ L_{l_B} \neq 0\ or\ L_{l_E} \neq 0 \\ \Delta L_l = \pm 1 \quad\ \ at\ L_{l_B} = 0\ or\ L_{l_E} = 0 \\ \Delta m = 0;\ \pm 1 \end{array}\right\} \qquad (7.103)$$

7.4.4.2. Optical emission spectra

The energy levels of the electron shells of multi-electron atoms have a very complex structure. Accordingly, their optical spectra are just as complex. For example, in the spectra of heavy atoms, a huge number of spectral lines appear, the number of which is measured in tens of thousands. However, the processes occurring in the absence of external excitations in the inner electron shells may not be considered, since they are mutually compensated and do not lead to any macro phenomena. This greatly simplifies the theoretical analysis of the optical spectra of multi-electron atoms.

The states of multi-electron atoms, as well as of the hydrogen atom (and all hydrogen-like atoms) will be denoted by the same, but with **capital** letters of the Latin alphabet. For example, the state of a hydrogen atom at $l = 0$ is denoted by the letter s, and the state of a multi-electron atom at $L_l = l_1 + l_2 + ... = 0$ is denoted by the letter S. The state of the hydrogen atom at $l = 1$ is denoted by the letter p, and the state of the multi-electron atom at $L_l = l_1 + l_2 + ... = 1$ is denoted by the letter P. Here l_1, l_2, l_3 ... are the orbital momenta of the outer electrons. The number of possible values of the total momentum, $N = 2\,L_l + 1$, and the **multiplets**, i.e., the number of lines into which the main line splits, is $t = 2L_S + 1$.

When the state of a multi-electron atom is designated, the main letter has two indices - the lower right indicates the total moment M, the upper left indicates the multiplet. Sometimes the principal quantum number of the atom is also indicated ahead of this designation.

Let us first consider the optical spectra of radiation arising from transitions of outer shell electrons on the example of two-electron atoms, i.e., atoms of helium and helium-like particles.

In the ground state of the helium (helium-like) atom $n = 1$, $l_1 = l_2 = l = 0$, and the spins of both electrons are antiparallel according to the Pauli exclusion principle, i.e.

$$m_{S_1} = +\frac{1}{2}; \quad m_{S_2} = -\frac{1}{2}$$

Both electrons are at the 1s level.

Therefore,

$$L_S = m_{S_1} + m_{S_2} = +\frac{1}{2} - \frac{1}{2} = 0$$

$$L_l = l_1 + l_2 = 0,$$

so

$$M = 0;\ t = 1.$$

The ground state of the atom corresponds, therefore, as we can see, to the unsplit (single, **singlet**) level.

This condition can be summarized as follows

$$1\ ^1S_0.$$

In the excited state, one electron is usually in the ground state 1s, and the second electron goes to one of the possible excited states. Since both electrons are in different states, their spins can be either antiparallel or parallel. In the first case $L_S = 0$, and in the second case $L_S = 1$. These states are shown in Figure 7.18 (a and b) respectively.

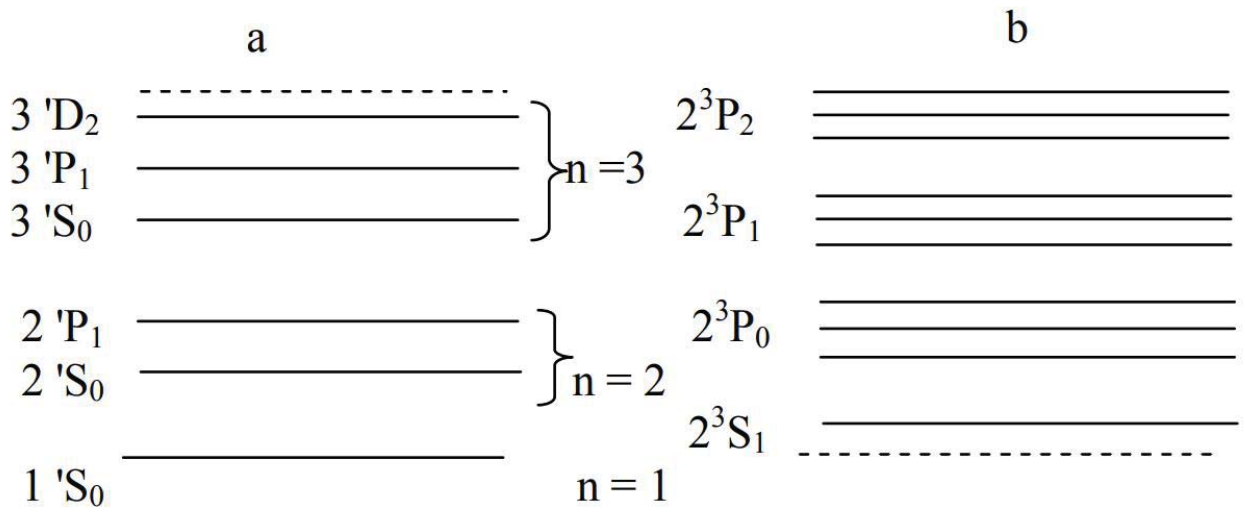


Figure 7.18.

If one of the electrons of an atom is not at the 1S level, its spin can be oriented either parallel or anti-parallel to the spin of the second electron. At antiparallel orientation, as in the previous case $S = 0$, the level is single (singlet) (see Figure 7.18)

The first state (see Figure 7.18a) is called **parahelium**, the second (see Figure 7.18b) is called **orthohelium**. Possible states of excitement of the parahelium will be:

$$2\ {}^1S_0\ (n = 2;\ l_1 = 0;\ l_2 = 0;\ L_1 = 0,\ L_S = 0;\ M = 0);$$

$$2\ {}^1P_1\ (n = 2;\ l_1 = 0;\ l_2 = 1;\ L_1 = 1;\ L_S = 0;\ M = 1)$$

$$3\ {}^1S_0\ (n = 3;\ l_1 = 0;\ l_2 = 0;\ L_1 = 0,\ L_S = 0;\ M = 0);$$

$$3\ {}^1P_1\ (n = 3;\ l_1 = 0;\ l_2 = 1;\ L_1 = 1;\ L_S = 0;\ M = 1)$$

$$3\ {}^1D_2\ (n = 3;\ l_1 = 0;\ l_2 = 2;\ L_1 = 2;\ L_S = 0;\ M = 2)\ \text{и т.д.}$$

All parahelium levels are singlet, because $t = 1$ for them. For orthohelium, the spins of both electrons are parallel, the spin momentum is equal to $S' = \frac{1}{2} + \frac{1}{2} = 1$, and $t = 3$, i.e. all lines form triplets. In

orthohelium, the second electron cannot be in the 1s state, since the first electron is already there.

Therefore, according to the Pauli principle, the state $1\ ^3S_1$ when $S' = 1$; $L = 0$, and $M = 1$ is excluded.

If the second electron is at the 2s level, the $2\ ^3S_1$ state occurs, which has the lowest possible energy and corresponds to the ground state of orthohelium.

If the second electron is at the 2p level, then $L = 1$ and $M = 0, 1, 2$. Three triplet levels correspond to this state: $2\ ^3P_0$; $2\ ^3P_1$, and $2\ ^3P_2$, etc. (see Figure 7.18b).

The spectra of atoms are also influenced by external fields. For example, the superposition of external electric and magnetic fields leads to a shift and splitting of spectral lines of energy levels, the appearance of a thin and hyperfine structure.

In this case, the states characterized by the same quantum numbers may differ in their magnetic moment projections to the direction of the magnetic field.

Optical spectra have very important practical applications. Qualitative and quantitative spectral analyses are the most important among them. Qualitative spectral analysis makes it possible to determine with high precision the chemical composition of the substance, the presence and chemical composition of even minor inclusions. With qualitative analysis, minor inclusions up to 10^{-13} kg can be detected. Quantitative analysis allows us to determine the concentration of the substance under study by the brightness of the spectral lines of its atoms. The sensitivity of the method is quite high.

Observation of spectra and spectral analysis is performed in the gaseous (glowing) state of the substance, in which there are no intermolecular and interatomic bonds.

7.4.4.3. X-ray spectra

X-rays are produced by bombarding solids with a stream of fast electrons in a vacuum X-ray tube. Colliding with the anode, which is made of a working substance, the electrons are dramatically decelerated and the resulting energy excites the electrons of the inner shells of the target atoms, which leads to their emission. Since the energy of the inhibited decelerated has a continuous distribution in a given interval, the spectrum of X-rays is continuous. The frequencies of this radiation are determined from the ratio

$$\mathcal{E}_1 - \mathcal{E}_2 = \hbar\upsilon, \tag{7.104}$$

where $\mathcal{E}_1$ is the initial energy of the electrons;

$\mathcal{E}_2$ is the energy arising in the process of deceleration of electrons.

Linear spectra of the so-called characteristic rays also appear on the background of the solid spectrum of the braking radiation.

It appears as a result of transitions of upper electrons to vacated internal levels.

Accordingly, *K*, *L*, *M*, etc. series of characteristic radiation can be distinguished.

Spin-orbital moments and their interactions are small compared to the huge excitation energy of the characteristic radiation and, therefore, there is almost no fine structure of the radiation here.

As early as 1913, Moseley established an empirical law linking the frequency of the lines of characteristic radiation with the charge of the nucleus *Z*, according to which

$$\sqrt{\upsilon} = c(Z - \sigma) \tag{7.105}$$

where c and σ are empirical constants.

Moseley's Law, in particular, is used in X-ray analysis. It is important to note that the results of this analysis do not depend on the

aggregate state of the substance, since the intermolecular interactions are mainly concentrated in the upper electron shells.

7.4.5 Fundamentals of quantum radiation theory

The classical theory of electromagnetic wave radiation is sufficiently described in Maxwell's electromagnetic field theory. In particular, it proves that the electromagnetic field has the property of self-sustaining and exists independently of the sources that excited it (see Sections 4.2.3 and 6.5.7). According to quantum theory, it is believed that radiation is caused by photons arising from the transition of a quantum system from an excited state with high energy to an unexcited state with less energy. In the reverse process, photons are absorbed and disappear because they cannot exist at rest, where their mass is zero.

Problems related to emission or absorption of photons in quantum theory are solved using the Schrödinger equation. However, the exact solution of this equation is possible only for the simplest systems. In the study of real-world problems, approximate methods are used. The most common among them is the so-called **perturbation method**.

This method consists in solving the Schrödinger equation for an unperturbed system in the first step. In the second step, the corrections due to small perturbations are calculated. Calculations of changes in quantum states show that the probability of transitions that cause radiation is proportional to the energy density of the electric field. The probability maximum arises at

$$\omega = \omega_{mn}, \tag{7.106}$$

which has the character of resonance.

Thus, an electromagnetic wave causes an electron to move from level m to level n at a frequency of

$$\omega = \omega_{mn}, \tag{7.107}$$

for which the value of

$$\omega_{mn} \approx \frac{U_m - U_n}{\hbar} \tag{7.108}$$

The probability of the reverse transition is equal to the probability of the forward transition, i.e.

$$p_{mn} = p_{nm}, \tag{7.109}$$

The calculated formula for the specified probability, obtained by the perturbation method, is (we will accept it without proof)

$$p_{mn} = \frac{E_0^2 e^2 |z_{mn}|^2 \sin^2 \frac{\omega_{mn} - \omega}{2} t}{\hbar^2 (\omega_{mn} - \omega)^2} \tag{7.110}$$

Where E_0 is the amplitude of the intensity of a plane monochromatic wave linearly polarized along the z-axis with length λ and frequency ω;

e is the value of the elementary charge;

$e|z_{mn}|$ is the average dipole moment of transition from *m* to *n*.

Forced transitions can occur both by absorbing the energy of the above field with its transition to a higher energy state and by giving energy to the field with its transition to a lower energy state.

Let a volume have a large number of atoms forming a rarefied gas, and the number of atoms at level *m* is N_m. The interaction of the atoms of this gas can be neglected. Let the gas at temperature T be in equilibrium with thermal radiation. Let us denote the number of atomic transitions from state *m* to state *n* by N_{mn}. If we denote the probability of such transitions per unit time by p_{mn}, then the number of transitions N_{mn} in time *dt* will obviously be equal to

$$dN_{mn} = p_{mn} N_m \, dt \tag{7.111}$$

Under the influence of an external electromagnetic field, induced transitions from a lower energy state to a higher state can also occur. Number of induced transitions dN mn in time dt is

$$dN_{mn} = p'_{mn} N_m u_{\omega 1T} dt, \quad (7.112)$$

where p'_{mn} is the probability of occurrence of induced transitions per unit time;

N_n is the number of atoms in state n;

$u_{\omega 1T}$ is the radiation density in dimensionless units.

The quantities N_{mn}, p_{mn}, and p_{nm} are called Einstein coefficients.

Einstein suggested that

$$p_{mn} = p_{nm} \quad (7.113)$$

The calculations show that in this case

$$\left.\begin{aligned} \frac{N_n}{N_m} &= \frac{e^{-\varepsilon_n/kT}}{e^{-\varepsilon_m/kT}} \\ U_{\omega_1 T} &= \frac{p_{mn}}{p'_{mn}} \frac{1}{\frac{p'_{nm}}{p_{nm}} e^{\hbar\omega/kT} - 1} \end{aligned}\right\} \quad (7.114)$$

Since spontaneous transitions occur from different atoms at different unrelated points in time, they cause incoherent radiation. Therefore, the radiation of all natural light sources such as stars, the Sun, an electric light bulb, a candle, a fire, etc. is incoherent.

In contrast to spontaneous radiation, induced radiation is coherent with the field that caused the radiation. Equality (7.113) implies mutual reversibility of induced absorption and emission transitions. And this is possible if the emitted photons are equivalent to the absorbed ones, i.e. photons of light that excite the radiation.

In the equilibrium state, the number of atoms in the normal (lower energy) state is usually greater than the number of radiating atoms. Consequently, the electromagnetic wave falling on the substance as it passes through the substance is attenuated (partially absorbed). It turned out, however, that in principle the inverse state of matter is possible, in

which the number of its excited atoms is greater than the number of atoms in the normal state. Such a state of matter is called **inverse**, inverted, reversed, or non-equilibrium. The process of putting matter into an inverse state, or, as they say, into a state with inverse population, is called **pumping**. At present, a number of pumping methods have been developed - thermal, optical, chemical, electro-ionization, etc. A medium filled with a substance with inverse population is called an **active medium**. In the active medium, the induced radiation exceeds the incident radiation. In other words, the light passing through the active substance is amplified and becomes coherent.

7.4.5.1. Quantum generators

The amplification of light by an active medium is used in quantum generators called **lasers**. The name is an acronym for "light amplification by stimulated emission of radiation". Lasers were created in 1954 by Soviet physicists Nikolay Basov and Aleksandr Prokhorov and, independently of them, by the American physicist Charles H. Townes.

The laser must have an active medium, a pumping system and an optical resonator. The optical resonator ensures that the laser generates strictly coherent radiation. Two facing each other and located on the same optical axis plane-parallel (or concave) mirrors (1 and 2), between which the active medium 3 is located (Fig. 7.19) are commonly used as a resonator.

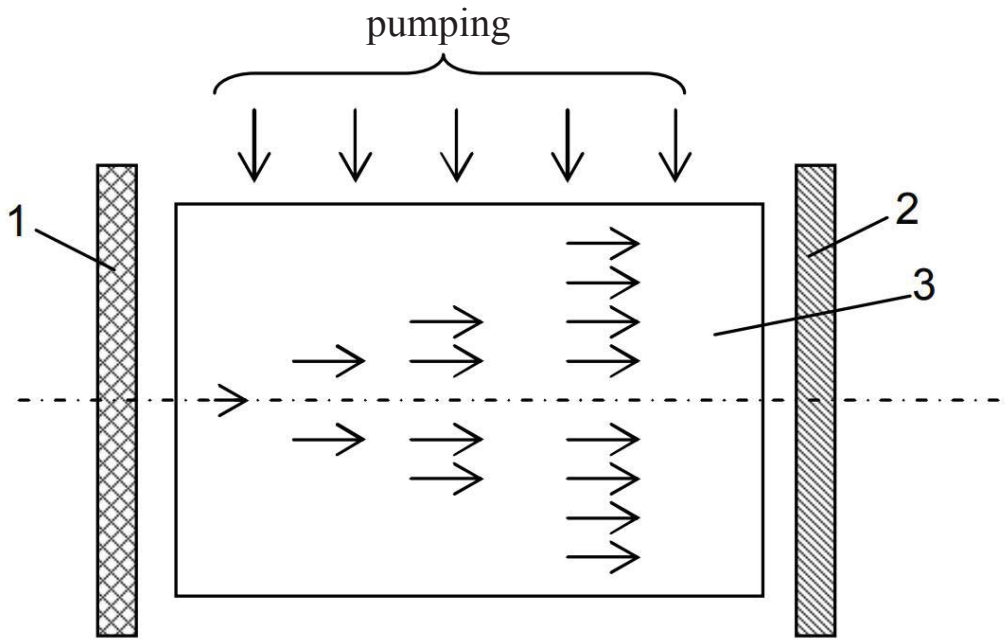


Figure 7.19.

One of the mirrors (2) must be partially transparent to allow radiation to escape outside the resonator. As a result of the spontaneous transition between the levels of the atoms of the active medium the excited atoms or ions accumulate at one or more of the so-called metastable levels, and then they are momentarily switched to the ground state, emitting light of the appropriate wavelength and phase. The phase shift of the radiation of different atoms is close to zero, and the waves emitted by different atoms are coherent.

The light wave (photon flux) emitted by each given atom, passing through the active medium, is amplified by repeatedly reflecting from the mirrors of the resonator. Photon fluxes, playing the role of induced radiation, cause secondary, then tertiary and so on radiation. Reaching the semi-transparent mirror 2, the radiation partly goes out and partly is reflected. The reflected radiation is used again for generation, forming a positive feedback. For the resonator to work properly, the reflected wave returning to a given point of the active medium must have a phase at that point that coincides with the phase of the primary wave.

In other words, (clause 6.6.1) there must be an integer number of half-waves on the resonator length L, i.e.

$$L = k\frac{\lambda}{2} \tag{7.115}$$

where L is the distance between the mirrors;

λ is the wavelength.

It follows from (7.115) that the frequency of modes generated by the resonator,

$$\omega_k = \frac{\pi k c}{Ln} \tag{7.116}$$

where n is the refractive index of the medium.

The modes that make up the laser radiation have different intensities in this case. The greatest intensity have, obviously, those modes whose frequencies coincide, i.e. resonate, with the frequency of radiation of inverse transitions. It should be kept in mind, however, that inverse transitions give an extended spectral line consisting of a whole set of frequencies. In this case, only one of the modes of laser radiation appears at the frequency of the main spectral line. The other frequencies giving low-intensity components in the composition of the laser radiation only slightly expand its spectrum. As a result, the spectrum of laser radiation is narrower than the spectrum of inverse radiation. In other words, laser radiation has a very high degree of monochromaticity. Considering that laser radiation in multimode mode includes waves (photons) differing not only in frequency, but also in phase, the mode close to single-mode is chosen, for which the dispersion of radiation in phase is minimal and, therefore, the best coherence is provided.

Let's take the ruby laser, one of the first quantum oscillators, as an example. The ruby laser consists of a ruby crystal in the form of a rod about 5 cm long, forming the active substance. From the chemical point of view, a ruby crystal is an aluminum oxide Al_2O_3, in the crystal lattice

of which some aluminum atoms (a few hundredths of a percent) are replaced by chromium ions, Cr^{+++}.

Figure 7.20 shows a scheme of induced optical transitions at laser pumping.

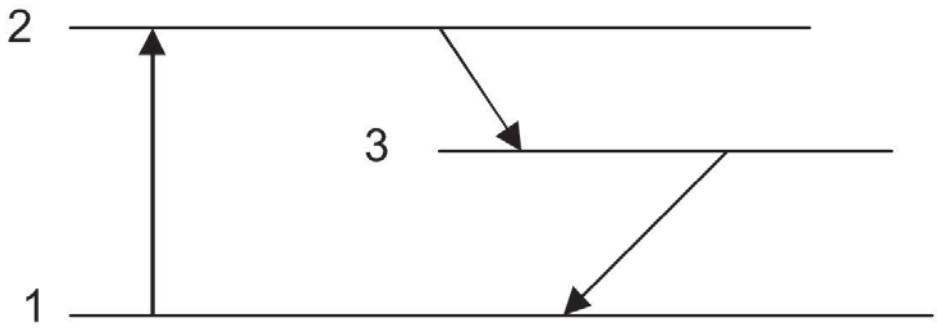


Figure 7.20.

Electrons of chromium atoms move from ground level 1 to excitation level 2 during the pumping process. The lifetime of atoms in the excited state is usually about 10^{-8} s. The electrons of chromium atoms do not return to the ground level (1), but to the auxiliary level (3), which is much closer to the ground level. Level 3 has the remarkable property that it is metastable. The chromium atoms' hold time at this level is several orders of magnitude longer (10^{-2} s) than at level 2. For this reason it is also called **radiation-free**. Spontaneous transition from level 3 to level 1 is prohibited by the selection rules. The energy released during this transition is transferred to the crystal lattice. As a result, significantly more excited chromium atoms accumulate at level 3 than at the initial level 1. In other words, there is an inversion of level 1 to level 3 in the crystal. The transition from level 3 to level 1 is proceeded as follows. First, under the action of external photons present in the matter, a few electrons move from level 1 to level 2. The resulting photons, reflected from mirror 1 (see Fig. 7.19), travel through the active medium, repeatedly and avalanche-like amplifying and causing avalanche-like induced coherent emission in the transition of electrons from level 3 to level 1.

Ruby lasers operate in pulsed mode. They are pumped by a high intensity green light. The optical resonator is created by carefully polishing the ends of the ruby crystal. One end is coated with an opaque layer of silver and the other with a thin translucent layer that allows about 8% of the incident radiation to pass through.

Gas lasers, in which a gas or gas mixture is the active medium, are also widespread. Gas lasers operate in a continuous mode. The inversion in gas lasers is carried out not by optical pumping but by a gas discharge. An example of a gas laser is the helium-neon laser.

Recently, preference has been given to semiconductor lasers in which the active medium is a extrinsic semiconductor, which is pumped by a direct current passing through the $p - n$ junction in the forward direction. Four-level gas lasers in which, in addition to the metastable level, a working level is also created between the metastable and the main level.

Lasers are widespread in many different fields of science, technology, medicine, and life. Laser technology is used for welding, cutting, melting metals, and drilling holes. In medicine, lasers are used for bloodless surgery and treatment of eyes, skin diseases, etc.

High coherence and monochromaticity of the laser beam is used in alignment, leveling works, in optical communication systems, in holography, as well as for the study of space objects. Laser spectroscopy, which allows observation of the spectra of individual particles, is also a promising technology.

Lasers are widely used in experimental physics to study high-speed processes, ultrafast photography. In everyday life, lasers are used in particular for playing audio-visual information, etc.

7.4.6. Elements of molecular quantum theory

As mentioned above, a molecule is the smallest particle of matter that has its basic chemical properties and consists of atoms connected by chemical bonds. The number of atoms that make up molecules ranges from two to hundreds of thousands in the so-called macromolecules that form polymers, including organic substances.

Atoms in molecules continuously make vibrational movements. Under certain conditions, for example in a gaseous environment, molecules can perform translational and rotational motion. Molecules, like atoms, have no clear boundaries or definite form. Conventionally, molecules are often considered to be spherical in shape. This approximation can be considered sufficiently true for simple molecules at distances sufficiently large compared to their sizes. If we assume that a molecule has a spherical shape, then, knowing the density of the substance, we can estimate the size of its molecules through the radius of the sphere.

The size of molecules grows as the number of atoms in them increases. Their shape changes accordingly. The approximate sizes of molecules are in the range $10^{-10} – 10^{-7}$ m, the shape of macromolecules can be linear, spiral, spherical, globular, etc. Macromolecules can be seen with modern electron microscopes.

Combining atoms into molecules is usually an energetically advantageous process, since the energy of a system of atoms in a bound state is less than in a free state. This, as mentioned above, is a consequence of the fact that the mass of a system of bound particles is always less than the total mass of these particles in the free state. The specified defect of masses of bound particles arises due to the fact that the lag, or, the same thing, inertia, and, therefore, mass as a measure of inertia, decreases with the increase of bonds and the corresponding convergence of particles during the transfer of interactions. On the other

hand, from the equivalence relation between mass and energy (see), the magnitude of the mass defect corresponds to a corresponding decrease in the energy of the system. This energy is transferred by the system to the surrounding bodies. Thus, in chemical fusion reactions, excess energy is generated and released in the form of heat.

Among the reactions of synthesis at the atomic level, which form a large number of bonds and are accompanied in this connection by the release of large amounts of heat, we should note the reactions of bonding with oxygen of carbon and many hydrocarbon compounds. These include, for example, hard coal, wood, gas, oil, petroleum products, peat, etc. These reactions are commonly referred to as combustion, and the materials listed are called combustibles or fuels. During combustion, a very large amount of energy is released, heating the burning substance (fuel) to such a high temperature that it transforms into a glowing plasma (fire).

To make the fusion reaction possible, energy must be expended to overcome the electromagnetic repulsion interactions of the atoms as they approach each other. This energy can be either less or more than the mass defect. In addition, bonding reactions are often accompanied by decomposition reactions, which go, on the contrary, with the absorption of energy. For these reasons, all chemical reactions are divided into reactions that take place with the expenditure of energy and reactions that take place with the release of energy.

A molecule is a quantum system, and therefore its internal energy is quantized. Roughly, we can assume that it is the sum of the energies of electronic movements $\mathcal{E}_{el}$, the vibrations of atomic nuclei $\mathcal{E}_{osc}$, and the rotation of the molecule as a whole $\mathcal{E}_{rot}$.

With this

$$\mathcal{E}_{el} >> \mathcal{E}_{osc} >> \mathcal{E}_{rot} \tag{7.117}$$

The state of a molecule is described by the Schrödinger equation, which decomposes into equations for electrons and for nuclei.

A molecule is an electrically neutral system, but its electron density is not uniformly distributed. The electronic levels of the atoms of the molecule are also superimposed on the interatomic interactions. Under their influence, each atomic level splits into many sublevels. The electronic levels of a molecule are determined by a set of quantum numbers characterizing the state of all its electrons, namely, total orbital momentum $\vec{L}_l$, total spin momentum $\vec{L}_S$, multiplet $t = 2\,L_S + 1$, and total momentum $\vec{M} = \vec{L}_l + \vec{L}_S$. The designation of the states of a molecule is the same as for a multielectron atom. The Schrödinger equation for nuclei describes vibrational and rotational levels. The oscillations of the nuclei are non-harmonic and are therefore represented as certain sums of harmonic oscillations.

A bi-atomic molecule can be viewed as a harmonic oscillator with frequency ω. Quantization of its energy gives

$$\mathcal{E}_{osc} = \hbar\omega\left(\chi + \frac{1}{2}\right) \tag{7.118}$$

where χ is the oscillatory quantum number;

A polyatomic molecule can be represented as a set of oscillators, so that

$$\mathcal{E}_{mol} = \sum_{i=1}^{f} \hbar\omega_i\left(\chi_i + \frac{1}{2}\right). \tag{7.119}$$

Rotational levels can be found by quantizing the rotational energy. For a two-atom molecule

$$\mathcal{E}_{rot} = \frac{M^2}{2I} \tag{7.120}$$

where M is the torque;

I is the moment of inertia relative to the axis of rotation perpendicular to the axis of symmetry of the molecule.

According to the rules of quantization

$$M^2 = \frac{\hbar^2}{4\pi^2} J(J+1) \tag{7.121}$$

where $J = 0, 1, 2, \ldots$ is the rotational quantum number.

Molecular spectra of different types occur at different transitions between energy levels

$$\Delta\mathcal{E} = \Delta\mathcal{E}_{el} + \Delta\mathcal{E}_{osc} + \Delta\mathcal{E}_{rot} \tag{7.122}$$

If $\Delta\mathcal{E}_{el} \neq 0$, the spectral emission lines of the molecules are observed in the visible and ultraviolet part of the spectrum. Rotational spectra are in the region with a wavelength of 0.1 - 1 mm, and vibrational spectra lie in the infrared region in the wavelength range of 1 - 50 mm.

Molecules are formed by combining atoms by means of so-called chemical bonds, which are created by electromagnetic interactions of external (valence) electrons of atoms weakly bound to their nucleus.

There are four types of chemical bonds - ionic, Van Der Waals, covalent, and metallic. An ionic bond is formed when the reacting atoms ionize, that is, when one atom gives up its outer electrons and the other atom attaches them. This happens most often when metals react with nonmetals. For example, the Na atom contains one electron in the M-layer and the 3s shell. It is shielded by two electron shells lying closer to the nucleus and is very weakly bound to the nucleus. The Cl atom, on the other hand, has a high ionization potential. Therefore, when the two atoms come within 0.25 nm ($2.5 \cdot 10^{-10}$) of each other, the outer electron of the Na atom goes to the Cl atom. The resulting Na^+ and Cl^- ions are mutually attracted and form a NaCl molecule. The newly formed electron shells of Na and Cl become completely filled with 8-electron shells.

All inorganic acids, alkalis, and salts (the connection of metals to acidic residues) are connected by the ionic bonding.

The van der Waals bond binds electrically neutral atoms without a dipole moment into molecules. Let us consider, for example, two hydrogen atoms that are so far apart that their electron clouds do not overlap. In the 1s state, the electron clouds are spherically symmetric, and the average dipole moment of these atoms is zero (the positive and negative charge of the nucleus and the negative charge of the electrons are combined). However, the instantaneous values of the dipole moments of each atom are not zero.

As atoms come closer to each other, their momentary dipole moments are either attracted or repelled. In the case of coordinated motion of the electrons, in which a constant attraction of the atoms is ensured, they form a stable molecule.

The interaction force of such atoms is called dispersion force. It has a quantum character and is explained by a decrease in the energy of interacting atoms. Van der Waals bonding in the case of polar atoms is orientational in nature (see Figure 7.21). Molecules with a high polarizability can also have an induced electric moment, creating an inductive attraction interaction. In practice, all of the above mechanisms of van der Waals interaction arise simultaneously. Despite this, they are very weak and are much closer to physical intermolecular interaction than to chemical interaction.

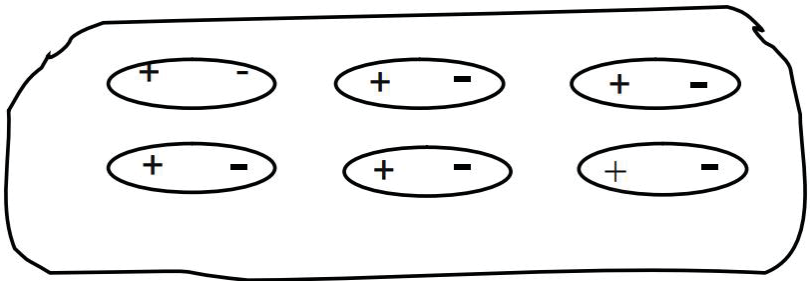


Figure 7.21.

The third type of chemical bonding combines identical atoms (e.g., O_2, N_2, etc.) or atoms with high chemical affinity into molecules. These include, for example, hydrogen-carbon, hydrogen-oxygen, nitrogen-oxygen, oxygen-carbon, and many similar pairs. Covalent bonding is leading in the formation of macromolecular polymers and all organic substances.

Covalent bonding is characterized by the collectivization of electrons. This happens when atoms get so close to each other that the electron clouds of their outer electrons overlap, and these electrons belong not only to their own nucleus, but also to the nucleus of the interacting atom (see Figure 7.22).

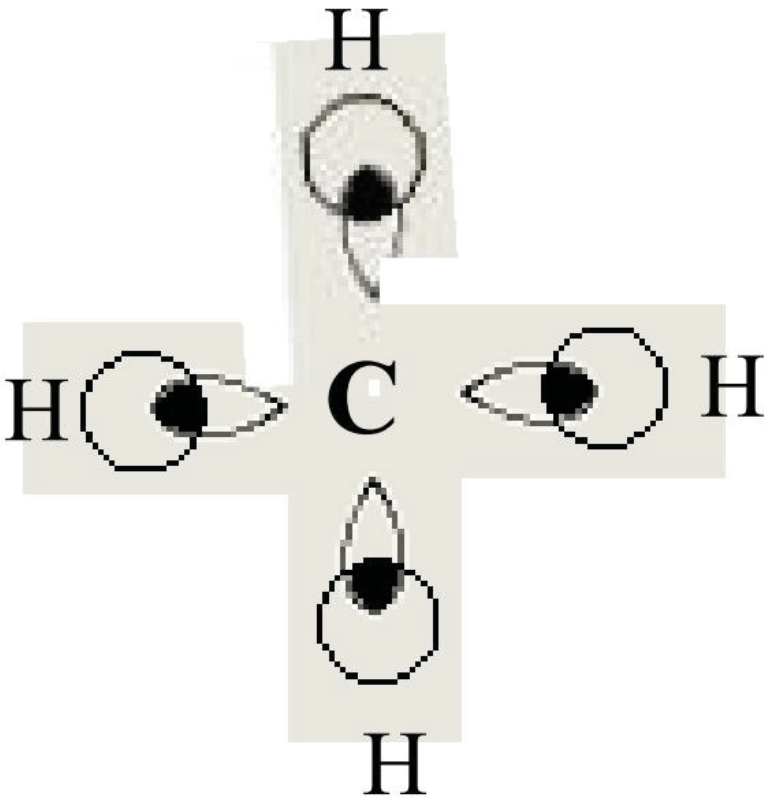


Figure 7.22.

In the methane molecule shown in Figure 7.22, the cloud of each s-electron of the hydrogen atom is superimposed on the clouds of the dumbbell form of the 2p-states of the carbon atom.

The fourth type, metallic bonding, is realized mainly in metal crystals. Each monocrystal can only conventionally be called a molecular aggregate. In fact, it is a giant molecule with a periodic structure, the so-called crystal lattice, in the nodes of which the positively charged ions of a given metal are located. These ions are formed due to the fact that the electrons very weakly bound to their nuclei bounce off them and create an electron gas that fills the cavities of the crystal lattice. The bonding between metal atoms is the result of the interaction between lattice ions and electron gas. This, by the way, implies a **long-range order of interaction**, when each ion interacts not only with the surrounding neighboring ions, but also with all the ions of the crystal, so that the entire crystal is a single entity. Thus, in the case of metal bonding, the electrons of all atoms of the crystal are collectivized.

In quantum theory, the geometry of molecules is characterized by a set of internal parameters - bond lengths, as well as the value of valence and torsion angles. The valence angles are the angles between two chemical bonds coming out of one atom, and the torsion angles are the dihedral angles of rotation around the bonds. For example, the hydrogen peroxide molecule H_2O_2 is represented as follows

H – O – O – H

The bond lengths of O – O and O – H atoms are 0.147 nm and 0.095 nm, respectively. The H – O – O valence angle is 95º. The torsional angle of rotation around the O – O bond, equal to 102º, is formed by the dihedral angle between the planes H – O – O and O – O – H.

Each state of a molecule has a characteristic equilibrium geometry, which is provided by its energy.

The symmetry of a molecule determines the symmetry of its wave function and is reflected, one way or another, in its properties - spectra, polarization in an external electric field, magnetic properties, geometry, etc. In particular, it leads to degeneration of electronic, vibrational and

rotational levels, determines exclusion of certain transitions, reduces the chemical activity of molecules. Molecules with a center of symmetry do not have a constant dipole moment. Molecules that do not have, in addition, a plane of symmetry form chemically active substances, etc.

Some molecules of the same composition differ in the order of their atoms and are called **isomers**. Isomers form substances with different physical and chemical properties. In addition to structural isomerism, there are also rotational (rotomers, conformers) and optical isomerism. The former arise from the rotation of atoms and their groups around chemical bonds and represent different states of the same molecule. The latter are mirror-symmetric molecules that rotate the polarization planes in opposite directions. These include, in particular, molecules of a number of organic substances that are part of living organisms (molecules of optically active substances).

7.4.7. Elements of quantum solid-state physics

The entire variety of properties of physical bodies is determined by the interactions of their electrons and atomic nuclei. Physical bodies in the solid aggregate state form crystalline structures. The only exceptions are so-called supercooled liquids or amorphous bodies, such as glass, wax, paraffin, etc.

The word "crystal" comes from the Greek word κρύσταλλος of the same sound (krústallos), which means "clear ice". Nowadays, crystals are called solids with a three-dimensional atomic periodic structure having the shape of a regular polygon. These include a regular pyramid, parallelepiped, tetrahedron, prism, dipyramid, octahedron, etc. A distinction is made between monocrystals and polycrystals. Monocrystals of table salt, for example, under equilibrium conditions of formation, repeat the shape of the elementary (i.e. the smallest) cell of the crystalline body. Polycrystals are bodies consisting of many chaotically oriented crystals of very small size, grains. All metals except liquid mercury are

typical polycrystals. All crystals have a cellular structure. Each cell has the shape of a regular polyhedron, which is a characteristic of this crystal. Numerous equally oriented cells form a crystal lattice. The parameters of the unit cell of a crystal lattice are determined by the way the atoms are arranged in it, and the properties of crystals as a whole are a function of the above types of chemical bonding between atoms. Ionic bonding is most often characteristic of dielectrics, which, when dissolved in water, dissociate into their constituent ions and form electrical current conductors of the second kind. Crystals with metallic bonds form conductors of the first kind. The electrical conductivity and thermal conductivity of metals are determined by the number of electrons on the outer shell and the degree to which they are bonded to the nuclei, which are chemically characterized by valence. Covalent bonds give crystals high hardness, low electrical conductivity, and high refractive indices (diamonds, for example). Van der Waals weak bonds are characteristic of molecular crystals. They are easily fusible and have low mechanical characteristics.

An elementary cell can be built up in different ways. The edges of a cell are called periods. A unit cell can contain from one atom (chemical elements) to 100 atoms (chemical compounds) and even 10^3 to 10^6 atoms (biological crystals). The lattice period ranges from a few angstroms to 10^3 Å. Any atom in a given cell corresponds to a translationally equivalent atom in every other cell (translation is a shift transformation).

7.4.7.1. Elements of the Band theory

Quantum theory of the energy spectra of electrons in solids (crystals) is called the **band theory**. This theory is based on the one-electron approximation, which proceeds from the following assumptions:

1) the mass of the nuclei anchored in the nodes of a perfect crystal lattice is incommensurably large compared to the mass of the electrons of the electron gas, and therefore the nuclei are assumed immobile relative to each other and the electrons;

2) atomic nuclei form an electric field with a periodic potential $U(\vec{r})$, in which the electron moves so that

$$U(\vec{r} + \vec{a}_n) = U(\vec{r}) \tag{7.123}$$

where $\vec{a}_n$ is the vector of the nth node defining the translation.

Based on this model, the Bloch theorem was formulated, according to which the wave function of the electron $\psi_k(\vec{r})$ in a crystal lattice

$$\psi_k(\vec{r}) = U_k(\vec{r})e^{j(\vec{k},\vec{r})} \tag{7.124}$$

where $\vec{k}$ is the wave vector of the electron.

In other words, according to Bloch's theory, the wave function of an electron moving in a crystal lattice coincides with the wave function of a free electron to an accuracy of the lattice period.

The wave function (7.124) can in this case be regarded as a solution of the special Schrödinger equation. It follows from this solution that the energy spectrum has the form of a series of allowed energy **bands** (**electronic bands**) $\mathcal{E}_l$ (k). Here l is the number of the allowed band. Allowed bands are obviously divided by band gaps.

It also follows from (7.124) that

$$\mathcal{E}_l\left(\vec{k} + \vec{b}\right) = \mathcal{E}_l\left(\vec{b}\right) \tag{7.125}$$

where $\vec{b}$ is the vector of the inverse lattice. The concept of the inverse lattice is abstract and is introduced into theory for the convenience of calculations. An inverse lattice is a lattice with lengths inverse to those of the forward lattice.

Since the wave vector $\vec{k}$ is expressed in terms of momentum $\vec{P} = \hbar\vec{k}$, the energy levels $\mathcal{E}_l\left(\vec{k}\right)$ can be viewed as functions $\mathcal{E}_l\left(\vec{P}\right)$, where $\vec{P}$ is a quasimomentum, different from the momentum of a free electron. This means that the electron in the crystal is some kind of quasi-particle

(quasi-electron) whose characteristics are called **effective** and are denoted by *. For example, such is the effective mass of quasi-electrons m^*. The function $\mathcal{E}_l(\vec{P})$ describes the energy structure of each band and is called the **dispersion law**, and the momentum components P_x, P_y, and P_z are the generalized coordinates of the momentum phase space. (see Section 2.5).

If the generalized coordinates are fixed, then $\mathcal{E}(P_x)$, $\mathcal{E}(P_y)$, $\mathcal{E}(P_z)$ are phase lines belonging to planes $\mathcal{E}(P_i)$. They are called **dispersion curves**. The entire set of $\mathcal{E}(P_i)$ forms a set of dispersion curves characterizing the function $\mathcal{E}(\vec{P})$.

If any value of energy is fixed

$$\mathcal{E}_i(\vec{P}) = const \tag{7.126}$$

then (7.126) is the isoenergetic surface equation. The family of surfaces $\mathcal{E}_i(\vec{P})$ in this case is characterized by the law of dispersion.

Band structure can also be introduced without resorting to solving the Schrödinger equation. Let a crystal be formed by N atoms, each of which in the free state has a discrete electronic energy spectrum. When these atoms are combined into a crystal lattice, a huge molecule is formed in which the electrons of all atoms are combined into a single quantum-mechanical system. The energy levels of the atoms are combined into an energy band containing, taking into account the spin, 2N levels. This band of allowed levels $\mathcal{E}_i(\vec{P})$. If there are Z electrons per one atom, then the total number of levels in the crystal is 2NZ. We are talking about discrete but fairly close to each other levels. The lower bands, corresponding to the levels of internal electrons, are quite narrow and completely filled. They occupy the levels of the allowed bands, starting from the bottom. The physical properties of crystals are determined mainly by the upper bands containing electrons. The interval between the bottom of the uppermost, still containing electrons, band $\mathcal{E}_2$ and the ceiling of the

previous completely filled band $\mathcal{E}_1$, is called the **band gap** (see Figure 7.23). The width of the band gap

$$\Delta\mathcal{E}_3. = \mathcal{E}_2 - \mathcal{E}_1$$

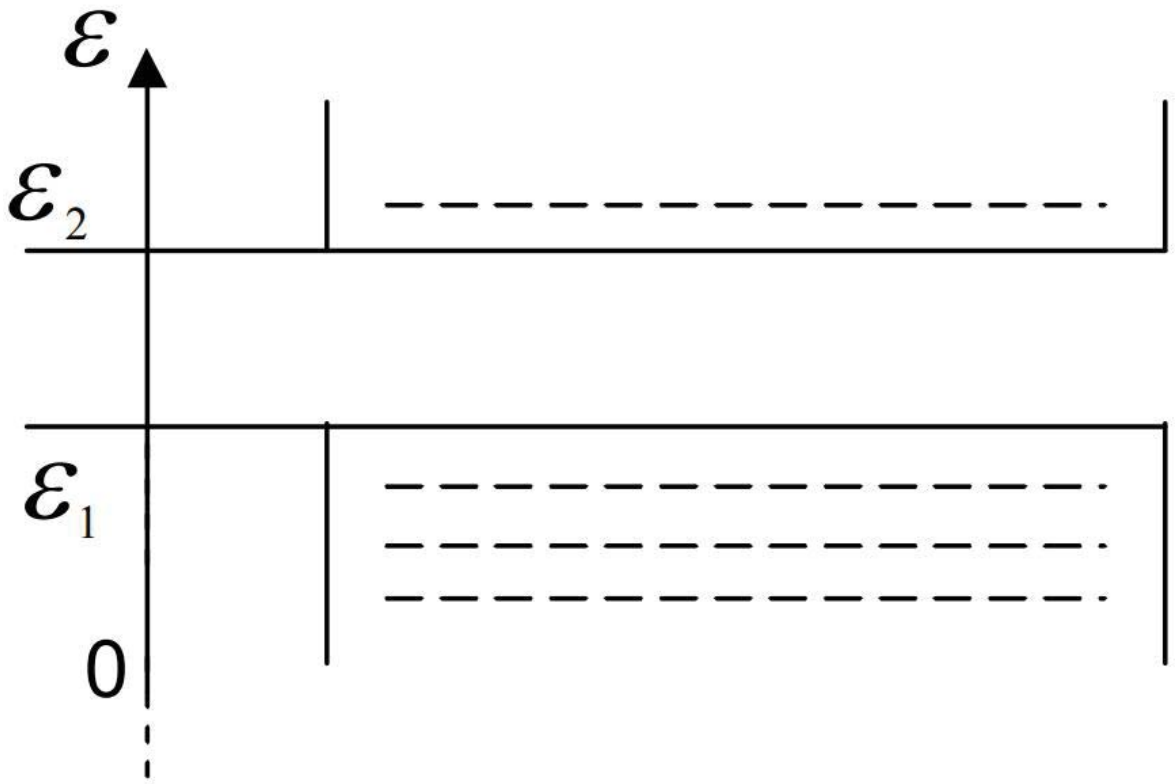


Figure 7.23.

Keep in mind that there may be several other allowed bands and band gaps below the filled band which are usually not considered.

If at temperature T = 0 all bands containing electrons are completely filled, the next upper band is therefore empty. It is separated from the nearest, fully filled band by a wide band gap. In this case, the crystal, in terms of electrical conductivity, is a dielectric. In metals, or 1st kind conductors, the band containing electrons is partially filled, the band gap is very narrow, zero or even negative (when the upper band containing electrons overlaps the fully filled lower band). Materials for which the band gap width is between that of dielectrics and metals are called semiconductors. For semiconductors, therefore,

$$\Delta\mathcal{E}_{met} < \Delta\mathcal{E}_{sc} < \Delta\mathcal{E}_{diel}$$

Materials with a perfect crystal lattice are called **pure**. The structure of real crystals always has some kind of disorder, which are called lattice defects. Defects most often occur during crystallization under the influence of mechanical, thermal, electrical, optical effects, as well as the introduction of impurities, etc. The simplest defects are vacancies (a lattice node in which an atom or its nucleus is missing), dislocations, which are lines along and near which the proper arrangement of atoms, characteristic of a crystal, is broken. There can also be foreign, doping agent atoms or ions present in crystal lattices, replacing the main atoms or embedding themselves between them. In addition, combinations of vacancies with electric current carriers are possible. Carriers of electricity in crystals, besides free electrons, can also be the so-called **holes**, which are quantum states in the atom that are not occupied by electrons.

Band theory made it possible for the first time to establish general patterns of electrical conductivity in all solids, conductors, dielectrics and semiconductors, and to obtain the exact mathematical relations for their description.

Semiconductors exhibit dielectric properties at absolute zero temperature. As the temperature increases, the semiconductor properties gradually prevail. Depending on the presence of doping agents, there are semiconductors with intrinsic and extrinsic conductivity. Semiconductors mainly include elements of groups 3 - 5, as well as elements of periods 3 and higher in the Periodic table. These are primarily Si, Gr, Ga, As, Se, Zr, and a number of others. In semiconductors with intrinsic conductivity, the bonds have a covalent nature, and each atom is associated with equidistant nearest neighboring atoms of 3 - 4 groups, the number of which is determined by its valence. Figure 7.24 shows the spatial lattice layer of silicon as an example. Since the silicon atom has 4 electrons on its outer shell, it forms covalent bonds with 4 neighboring atoms.

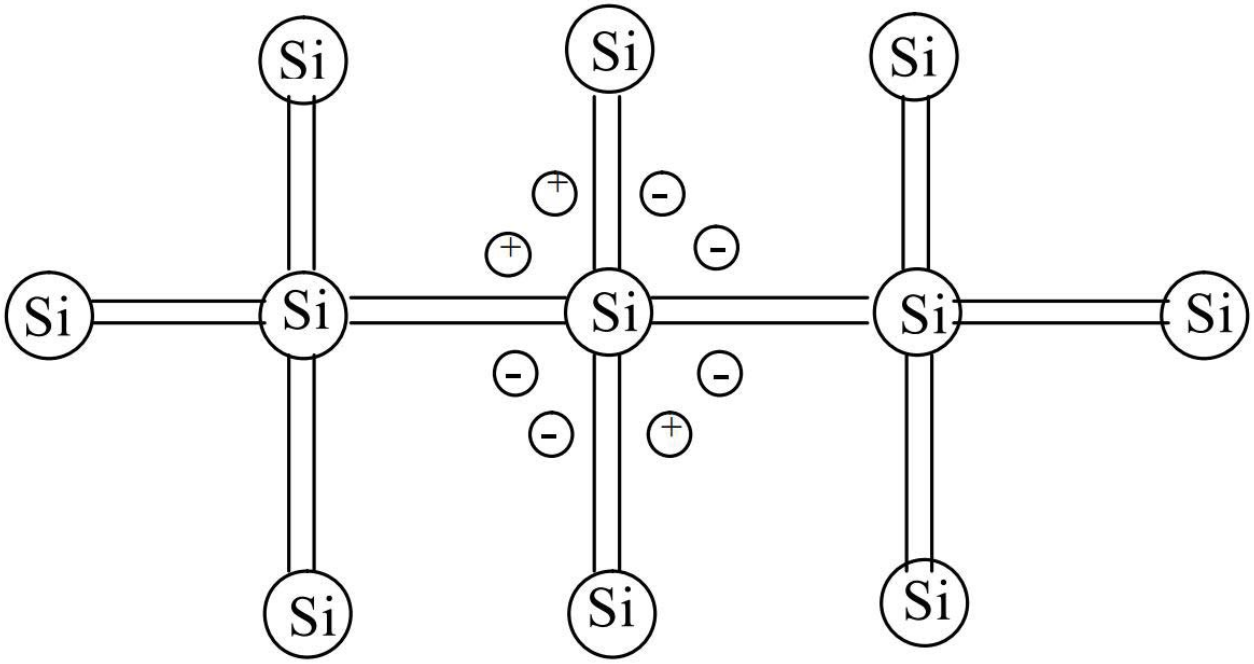


Figure 7.24

As a result, each pair of atoms formed by this atom with 4 neighboring atoms will have 2 electrons in common.

In some cases, one or both of the electrons of a collectivized pair are replaced by holes. In essence, this means that one or both electrons of the pair have detached from their nuclei and moved into intercellular space. This occurs as a result of breaking individual bonds under the influence of temperature or other factors.

When one bond is broken, one electron is released and a positive uncompensated charge in the form of a hole appears in the vicinity of the place it left. If the resulting hole recombines with another electron, it goes to the empty place it left. Under the action of a directed external field, the free electrons move in one direction and the holes in the other.

In terms of the band theory, when a semiconductor is excited, a valence electron moves from the filled (valence) band to the conduction band and becomes a free carrier of negative charge, leaving instead a hole in the filled band, which is also a free carrier, but with a positive charge.

During recombination of a free electron with a hole, the electron passes, on the contrary, from the conduction band to the valence band.

* Under the action of an external directional electric field in a semiconductor at temperature T ≠ 0, a current arises with a density of $\vec{j}$

$$\vec{j} = \vec{j}_e + \vec{j}_p \tag{7.127}$$

where $\vec{j}_e$ is the vector of electron current density;

$\vec{j}_p$ is the vector of hole current density.

From the definition of the current density vector it follows that

$$\left.\begin{aligned} \vec{j}_e &= n_e(-e)\vec{v}_{eav} \\ \vec{j}_p &= n_p e\vec{v}_{pav} \end{aligned}\right\} \tag{7.128}$$

where n_e and n_p are concentrations of electrons in the conduction band and holes in the valence band, respectively;

$\vec{v}_{eav}$, $\vec{v}_{pav}$ are the average speeds of directional motion of electrons and holes. With this

$$\left.\begin{aligned} \vec{v}_{eav} &= \Delta\tau_e\vec{a}_e \\ \vec{v}_{pav} &= \Delta\tau_p\vec{a}_p \end{aligned}\right\} \tag{7.129}$$

where $\Delta\tau_e$, $\Delta\tau_p$ are the free path times of carriers between two successive collisions with lattice ions;

$\vec{a}_e$, $\vec{a}_p$ are the accelerations of the directional motion of the carriers.

From the second law of dynamics is follows that

$$\left.\begin{aligned} \vec{a}_e &= \frac{e\vec{E}}{m_e *} \\ \vec{a}_p &= \frac{e\vec{E}}{m_p *} \end{aligned}\right\} \tag{7.130}$$

where $\vec{E}$ is the external field strength;

m_e*, m_p* are the effective masses of the carriers.

On the other hand, according to Ohm's law for specific conductivity γ

$$\vec{j} = \gamma \vec{E} \tag{7.131}$$

After appropriate substitutions we obtain

$$\gamma = e^2 \left(\frac{n_e \Delta\tau_e}{m_e *} + \frac{n_p \Delta\tau_p}{m_p *} \right) \tag{7.132}$$

Since, by definition

$$n_e = n_p = n, \tag{7.133}$$

then finally

$$\gamma = en\,(k_e + k_p), \tag{7.134}$$

where k_e, k_p are the mobility of carriers –

$$k = \frac{\upsilon_{av}}{E} \tag{7.135}$$

The emergence of an electron-hole pair is called generation, and the opposite process of its disappearance is called recombination. It is easy to understand that the probabilities of generation and recombination in the stationary state are equal

$$p_1 = p_2.$$

These probabilities are known (see Section 7.2.1.2) to be determined from the relation

$$p_1 = e^{-\frac{\Delta W \partial}{kT}} \tag{7.136}$$

whence, since the recombination probability is proportional to the concentration of both electrons and holes, then

$$p2 = an_e n_p.$$

It follows from (7.133) that

$$\gamma = \gamma_0 e^{-\frac{\Delta W \partial}{2kT}} \quad (7.137)$$

In semiconductors with intrinsic conductivity, the number of electron and hole carriers is equal to each other. In semiconductors with extrinsic conductivity the situation is different. If a small number of 5-valent atoms are introduced into the crystal lattice of a 4-valent semiconductor, the 5th valence electron cannot form a covalent bond and is superfluous and already at a sufficiently low excitation energy becomes free, and the atom becomes a positively charged ion (forms a hole). As a result, conduction electrons of doping agent atoms appear in the semiconductor, and the atoms themselves form donor levels.

If 3-valent doping agent atoms are introduced into the crystal lattice of a 4-valent atom, then, as is easy to see, free charge carriers in the form of holes appear in the crystal, and the atoms themselves form acceptor levels.

From the point of view of the band theory, doping agent atoms are distributed chaotically throughout the crystal lattice of the main atoms and thus violate its periodicity. The energy levels of doping agent atoms are located in the band gap, since they are located in the intercellular space and are subjected to additional influence from the atoms of the crystal lattice. Their bonding energy, therefore, is lower than that of the main atoms, but higher than that of the free carriers. For this reason, their levels are above the ceiling of the valence band, but below the bottom of the conduction band. Extrinsic levels are called **local levels**. Their free electrons and holes are localized near their atoms and therefore are not free enough to move into the conduction band and create a conduction current under the action of an external field. With this the energy of free electrons of doping agent atoms is somewhat lower than that of the free electrons of the main atom, i.e., their levels are located near the bottom of the conduction band. For the same reason, the levels of free holes are located near the ceiling of the valence band.

Under the influence of an external field, electrons of donor levels move into the conduction band and create an electronic current, and holes of acceptor levels move into the valence band and create a hole current. Therefore, semiconductors of the first kind are called n-type or electron semiconductors, and those of the second kind are called p-type or hole semiconductors. (see also Section 4.1.4.6).

7.4.8. Quantum Liquid

According to the concepts of classical physics, the kinetic energy of bodies decreases in proportion to the decrease in temperature. At the temperature of absolute zero, the thermal vibrations of the particles must stop, i.e. they must take a certain position, and the body must go to the crystalline state.

Experience, however, shows that some bodies retain a liquid state or even acquire the property of superfluidity. In addition, most metals, while remaining solid, acquire the properties of superconductivity. This is explained by purely quantum effects. According to the uncertainty relations, the kinetic energy of particles at absolute zero temperature is not equal to zero. The particles in this case continue to make the so-called **zero-point energy fluctuations**. When the amplitude of the zero-point fluctuations is comparable to the average distance between the particles, the body remains **liquid** even at absolute zero temperature. For this, however, it is necessary that the intermolecular interaction forces between the particles and their masses be sufficiently small, as, for example, in helium isotopes. Such fluids are called **quantum fluids**.

Quantum liquids are divided into Bose–Einstein condensate and Fermi liquids, depending on the spin value of the particles that form them. For example, the spin of helium isotope 4 with an even number of nucleus particles is zero. This isotope belongs to Bose–Einstein condensate, and helium isotope 3, which has a spin of 1/2, belongs to Fermi liquids.

The electron gas formed by conduction electrons in the crystal lattice of metals is also referred to Fermi liquids.

The properties of quantum liquids were discovered and studied by the Soviet physicist Pyotr Kapitsa, and the theory of helium superfluidity was developed by another Soviet physicist Lev Landau.

The essence of superfluidity of quantum liquid is that when it flows through a thin tube or slot at a speed less than the critical one, it meets no **braking resistance** from the walls of the tube (slot). Quantum liquid formed by conduction electrons in the crystal lattice of metals contributes to the appearance of superconductivity of these metals. These properties of quantum liquid are explained as follows. According to quantum mechanics, any system of interacting particles can only be in certain states that are characteristic of the system as a whole. In this case the energy of the system can change only discretely, which is equivalent to the birth or disappearance of excitations, which can be considered as **elementary quasiparticles** with their own energy, momentum, spin, etc. Quasiparticles include, for example, the so-called **phonons**, which characterize the elastic excitations of the system. In Fermi liquids, quasiparticles arise and disappear only in pairs, which are sometimes called **Cooper pairs**.

At low temperatures, the number and interactions of quasiparticles are small, and the quantum liquid is close to an ideal gas. At absolute zero temperature, quasiparticles tend to occupy the state with the lowest energy, which, taking into account the Pauli exclusion principle, fills a certain, so-called **Fermian sphere** in the impulse space. Radius of the Fermian sphere outside which the concentration of particles $n = 0$,

$$P_F = (3\pi^2)^{1/3} n^{1/3} \hbar \qquad (7.138)$$

At temperatures other than zero, quasiparticles with momentum larger than the radius of the sphere appear. These quasiparticles are

outside the Fermian sphere, leaving instead inside the sphere, respectively - **quasi-holes**. The ratio of momentum to velocity calculated on a Fermian sphere is called the **effective mass** of the quasiparticle.

The interaction of quasi-particles at absolute zero temperature manifests itself as a sound wave propagating in quantum Fermi liquid, which is called **zero sound**.

If the interaction of quasiparticles of a Fermi liquid has an attraction character, then the so-called **Cooper pair** of electrons arises, which causes superconductivity. This phonon-exchanging pair is called a Cooper pair, after the scientist who discovered and measured the bonding energy between its electrons.

The properties of superfluid helium 3 are significantly different from those of superfluid helium 4 and from Fermi-liquid in superconductors.

7.4.9. Quantum theory of superconductivity

The phenomenon of superconductivity was first observed in 1911 by the Dutch physicist Heike Kamerlingh Onnes, who discovered that at a temperature of about 4.15 K the electrical resistance of mercury abruptly disappears. It turned out that many other metals and alloys have similar properties.

The property of conductors to **drop their resistance to zero by leaps and bounds** at a low, but certain, **critical temperature** is called **superconductivity**.

In addition to the critical temperature, superconductors are also characterized by the value of the critical magnetic field induction (see Figure 7.25).

If the superconductor is in an external magnetic field, it is displaced to its surface. This phenomenon is called the **Meissner effect** (or Meissner–Ochsenfeld effect).

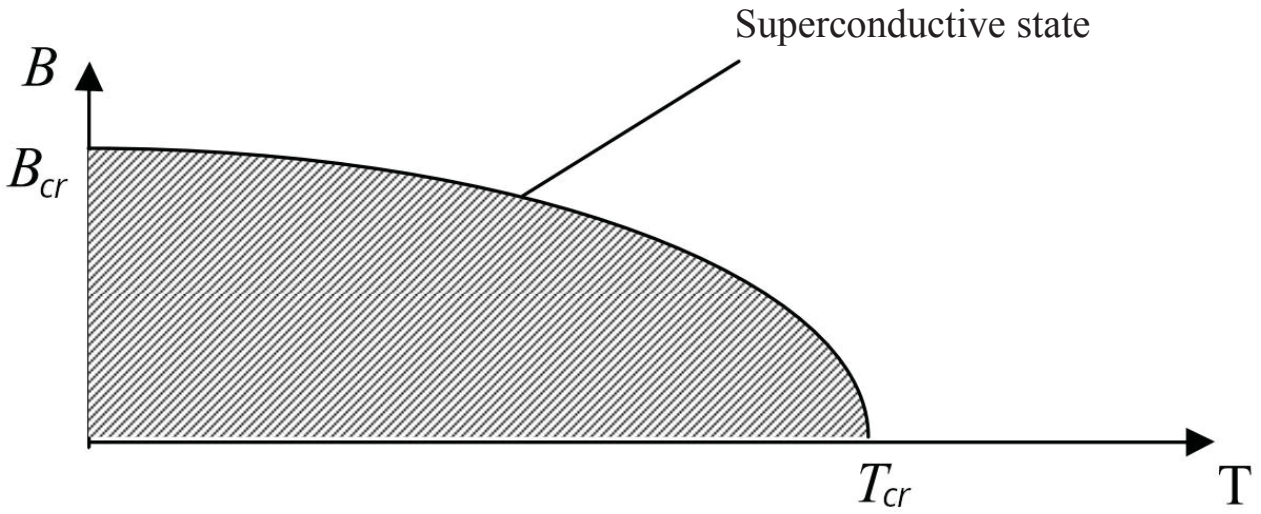


Figure 7.25

The undamped electric current in the superconducting state of the conductor is concentrated in its surface layer. The thickness of this layer should be such that the magnetic field created by it would exactly compensate the external field.

The thickness of the surface layer within which the magnetic field retains a non-zero value is called the **penetration depth**. The penetration depth is between 10^{-7} and 10^{-8} m.

Let us consider a metallic sample cooled to a temperature less than the critical temperature and thus being in a superconducting state. If we introduce this sample into an external magnetic field with induction $B < B_{cr}$, it will disappear inside the sample.

This is explained by the fact that the penetration of the field inside the sample leads to the appearance of electromagnetic induction electromotive force in it. This electromotive force induces an induction

current in the sample, the magnetic field of which is directed against the external field.

It follows from zero resistance of the sample that the internal field instantly increases to the value of the external field and thus completely compensates it.

This means that inside the sample the field induction $B = \mu\mu_0 H = 0$. It follows that $\mu = 0$. In other words, superconductors are ideal diamagnetic materials (see Section 4.2.2.6).

According to their behavior, all superconductors are divided into superconductors of the 1st and 2nd kind, which differ significantly in the magnetization curves M(B) of the unit volume moment dependence on the magnetic field induction (see Figure 7.26).

In the initial section, at small values of the magnetic field strength B up to the critical value of the field strength B_{cr}, the magnetization curves of superconductors, both 1st and 2nd kind (see Figures 7.26 a, 7.26 b), have a linear nature and practically do not differ from each other.

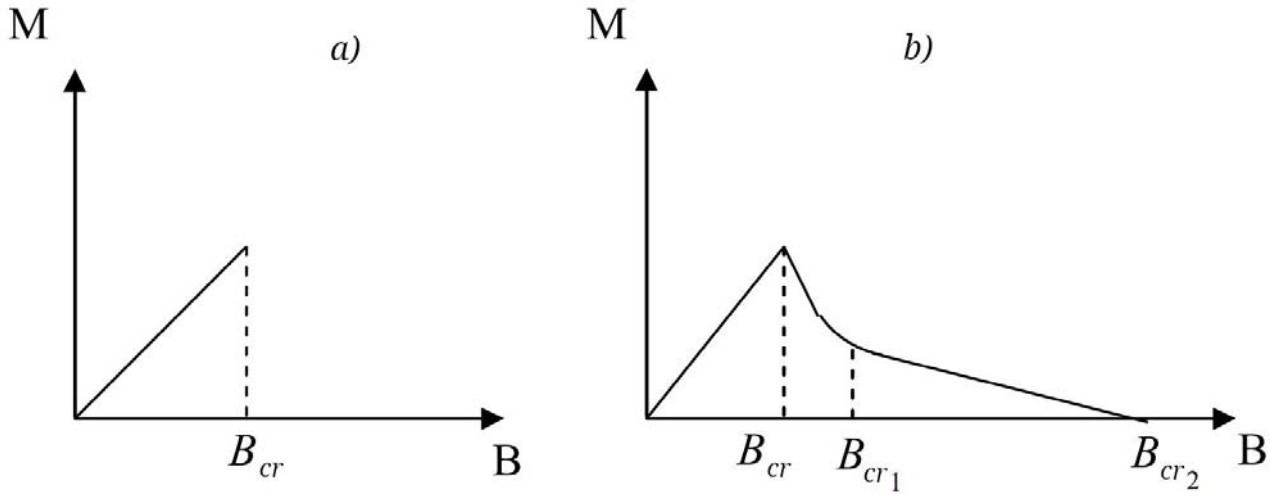


Figure 7.26.

When the field increases above the critical value, superconductors of the first kind go to the normal state, and the magnetic moment of the second kind decreases gradually, so that in the interval between the field

strength $B_{cr}1$ and $B_{cr}2$ they retain their superconducting properties. Since the resistance of the superconductor is zero, Ohm's law does not apply to it. This means that the current in the superconducting state of the conductor at a given voltage does not tend to infinity, but is limited by the amount of charge carried by the field per unit time.

The current intensity in this case is, of course, greater than in a normal conductor at the same voltage, but in the superconducting state it is not the current intensity that increases dramatically, but its density, because the current is displaced to the surface of the conductor. At the same time, due to the superconductor's lack of resistance, it does not heat up even at enormous current densities. Another important feature of the superconducting state is that the current in the unloaded conductor does not decay over time after the voltage is turned off. It should be borne in mind, however, that the magnitude of the current in a superconductor is limited by the value of the critical field, which is introduced into the conductor by the current.

Purely theoretically, all metals have superconducting properties, but they manifest themselves differently in different metals. Some metals, such as Cu, Ag, Au, Pt, Fe, Co, Ni, and most alkali metals have a critical temperature so close to absolute zero that it is impossible to observe their superconducting properties. Superconducting properties are strongest in metals of the rare-earth group, including Nb, V, Ga, Ti, etc. Pure metals are most often referred to as superconductors of the 1st kind. Superconducting properties have now also been found in several hundred alloys, although the components of these alloys themselves are not always superconductors. Most of the alloys belong to the superconductors of the 2nd kind. The critical temperatures of superconductors vary from very low, such as 0.01 K for tungsten, to very high, such as 23 K for the Nb3Ga alloy. If superconductivity occurs at critical temperatures greater than 25 K, it is called **high-temperature superconductivity**. Currently, alloys with high-temperature superconductivity up to 125 K have been discovered. This is, for example, the alloy $TiB_2Ca_2Cu_3O_{10}$. The

advantage of high-temperature superconductors is that relatively cheap coolants are used for their cooling instead of expensive liquid helium. For example, liquid hydrogen is used up to the critical temperature of 90 K, and liquid nitrogen for higher temperatures. In principle, the possibility of obtaining high-temperature superconductors with a temperature of 300 K and above, which can be cooled by flowing water is not excluded.

It is impossible to explain the phenomenon of superconductivity with the help of classical theory. This is due to the fact that, according to this theory, the resistance of a conductor can decrease to zero only at absolute zero temperature. In addition, for this it is necessary to assume that at absolute zero temperature all possible fluctuations of current carriers completely disappear. This is not really observed in the real world. As opposed to classical theory, quantum theory allows giving not one but several explanations of this phenomenon.

The most common one is based on the theory of Fermi gas, which is the basis for the superfluidity phenomenon of helium described by Landau. This also takes into account the interaction of conduction electrons between themselves and the crystal lattice of metal conductors.

The mechanism of this interaction is as follows. Free electrons are known to create an electron gas in the metal, which moves in the crystal lattice. This lattice is formed by positively charged ions fixed in the nodes of the lattice and making thermal vibrations relative to the position of equilibrium. These vibrations propagate in the crystal lattice in the form of elastic acoustic waves. The quantization of these waves produces a set of sound field quanta, phonons. Since ions are charged, their displacement from the equilibrium position is accompanied by polarization of the crystal. This polarization propagates as waves in the crystal lattice and acts with its electric field on the electrons of the electron gas. As a result, besides the usual Coulomb mutual repulsion forces, weak mutual attraction forces arise between conduction electrons,

which, under certain conditions, unite these electrons into stable and independent of each other Cooper pairs.

In an external electric field, the entire set of Cooper pairs moves in the direction of the field as a unit, without interacting on average with the crystal lattice, that is, without encountering any resistance. In the case where the number of outer shell electrons is odd (1 or 3), their Cooper pairs do not completely overlap the crystal lattice, and in the conductor may remain excited areas that destroy the pairs and the overall superconductivity of the conductor. This explains the fact that the metals of these groups, as mentioned above, usually do not exhibit superconducting properties even at very low, close to absolute zero, temperatures.

The mechanism of superconductivity is explained in quantum theory as follows. Let some electron 1 interact with the crystal lattice and put it in an excited state. The resulting polarization wave is equivalent to the emission of a phonon, which transfers some energy and momentum to the lattice. Electron 2, which is close to electron 1, interacts with the excited crystal lattice and takes on the energy and momentum of excitation, which is equivalent to its absorption of a phonon. Due to the exchange of electrons of 1 and 2 pairs, the resulting interaction of conduction electrons with the crystal lattice is compensated by the phonon. This means that on average, i.e. during the total time of emission and absorption of the phonon by pair 1 and 2, the resulting excitation of the system disappears, and the electron gas and the crystal lattice return to their initial state. The other pairs of electrons behave similarly, provided that each pair, exchanging phonons, is stably bound together and does not interact with other pairs.

The phonon exchange mechanism was discovered in 1950. In 1956, Cooper solved the quantum mechanical problem of the phonon attraction of electron pairs. He defined the phonon attraction binding energy 2Δ as a function of the lattice oscillation frequency ω_D (the

Debyean frequency of phonons) and the density of electronic states to Fermi levels υ ($\mathcal{E}_p$) through the electron-phonon interaction constant g

$$\Delta \approx \hbar\omega_D e^{-\frac{2}{g\upsilon(\mathcal{E}_p)}} \tag{7.139}$$

It was found that the stability of the pair is ensured only if the energy of thermal fluctuations does not exceed the bonding energy of these pairs.

Calculations show that the bonding energy of these pairs is relatively low and is in the range of $10^{-3} \cdot 10^{-4}$ eV. This energy corresponds to a small value of absolute temperature in the range (2 – 25) K. If this condition is met, the movement of pairs is rigidly coherent.

This means that Cooper pairs cannot change their states independently of each other. The coherence of these states also means a high degree of orderliness of the pairs' motion and the independence of this motion from the characteristics of the system and the pairs themselves.

The phenomenon of superconductivity leads to another interesting, so-called stationary **Josephson effect**. This effect is that if two metals are separated by a very thin layer of dielectric (on the order of 10^{-9} m), called a tunnel contact, then conduction electrons can flow through it, as through a potential barrier, and move from one metal to the other.

When these metals are in a superconducting state, Cooper pairs will leak through the tunnel contact. Under the action of the DC voltage applied to the contact, a superconducting current is generated in it. If this current does not exceed some critical value, determined by the critical value of the magnetic field strength, the voltage drop in the tunnel contact is zero.

When a current greater than the critical current is passed through the tunnel contact, a non-zero voltage drop occurs across the contact.

Under the action of this voltage, the contact emits monochromatic radiation.

This means that an alternating current of frequency $\omega = 2eU/\hbar$ flows through the contact to which a constant voltage is applied. Here, 2eU is the difference in the principal standing energies of the Cooper pairs in the left and right metals (see Figure 7.27a).

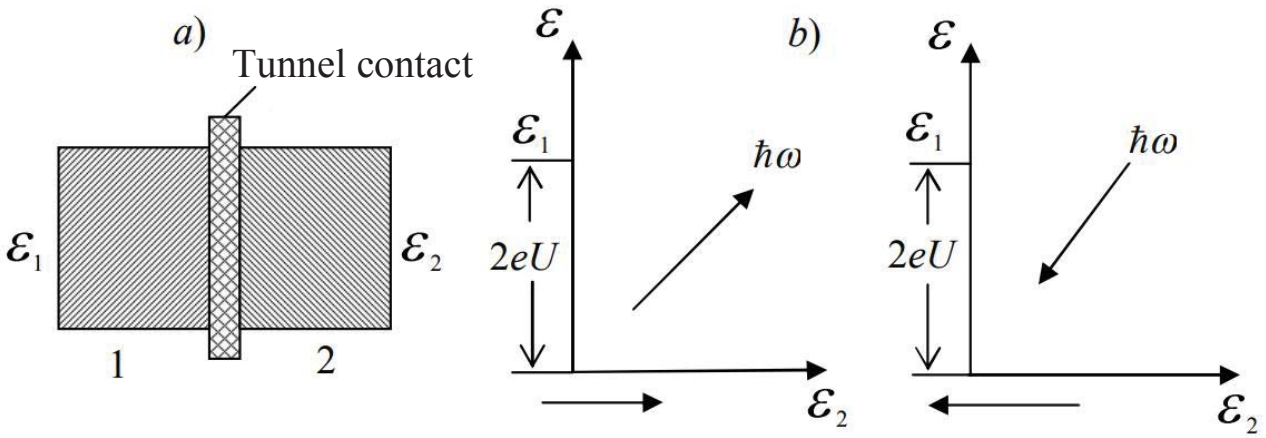


Figure 7.27.

It should be taken into account that one metal (1) radiates and the other metal (2) absorbs photons with energy $\hbar\omega$.

Monochromatic contact radiation is used in interferometers of the highest sensitivity and accuracy.

The device based on the Josephson stationary effect we discussed, used for high-precision measurement of weak magnetic fields, is called a **SQUID**. SQUID is an acronym for superconducting quantum interference device.

SQUIDs have also found applications in biology and medicine to measure the magnetic fields of weak currents associated with heart and brain function. In electronics, SQUIDs are used as high-speed computer elements.

7.4.10. Fundamentals of quantum theory of the atomic nucleus. Internal quantum numbers.

So far we have considered processes that are the consequence of electromagnetic interactions occurring at the atomic-molecular level. They also include intermolecular interactions that cause thermal phenomena and chemical processes.

Let us now turn to the processes that arise as a consequence of **intranuclear interactions**.

The atomic nucleus was discovered by the famous British physicist Rutherford in 1911 in experiments on the **scattering of α-particles**. Rutherford's hypothesis, according to which the positive charge of the atom is concentrated in its center (the nucleus) and occupies a minuscule volume (according to Rutherford's calculations it is 15 orders of magnitude below the volume of the atom as a whole), was confirmed first by the works of Bohr, and then by direct experiments of English physicist Moseley. Currently, the size of the nucleus of atoms is estimated to be in the order of 10^{-15} m.

It turned out that the atomic nucleus, apart from the nucleus of the hydrogen atom, in turn has a rather complex structure, the elements of which are connected by **strong** intranuclear interactions (see Section 1.2.2).

The quantum theory of the atomic nucleus is based on its universally accepted model, proposed in 1932 by the Soviet physicist Dmitri Ivanenko and the German physicist Werner Heisenberg. According to this model, the nucleus of atoms consists of electrically charged **protons** with a positive charge numerically equal to the electron charge ($1.6 \cdot 10^{-19}$ C), and uncharged **neutrons**. Only the nucleus of the hydrogen atom contains no neutrons and consists of only one proton. The proton charge is conventionally assumed to be 1. The charge of the nucleus Z, expressed in conventional units, determines the chemical

properties of atoms and their serial number in the Periodic table. The proton and neutron masses refer to **fundamental constants** and are approximately $1.67 \cdot 10^{-27}$ kg (938.4 MeV/s^2). Protons and neutrons are otherwise indistinguishable, so they are often considered to be one and the same particle, a **nucleon**, in different charge states. Nucleons are fermions and have a spin value equal to 1/2.

Proton is characterized by enormous stability. Its lifespan, according to modern concepts, is about 10^{31} years in the free state. The neutron in the free state is not stable. It lives only 15.3 minutes and decays into a proton, an electron, and an antineutrino. On the contrary, in the bound state, for example in the nuclei of stable atoms, neutrons are highly stable.

Nucleons belong to the group of massive intranuclear particles, hadrons, which are characterized by a specific baryonic charge B = +1 and are called **baryons**. Baryons, except protons, are generally low-stability particles. Antibaryon charge B = –1. Electrons, neutrinos, and antineutrinos are not part of the nucleus, but take an active part in intranuclear processes caused by weak interactions (see section 1.2.2). The sum of the number of protons Z and the number of neutrons N in the atomic nucleus is called the **mass number A**. Since all atoms without exception are electrically neutral, the number of protons in the nucleus Z is always equal to the number of electrons of the atom. The number of protons and neutrons in the light and middle atomic nuclei are approximately equal. As the mass number of the atomic nucleus increases, the electrical repulsion of protons increases, which is compensated by the increase in the specific neutron content in the atom. For example, the number of neutrons in the nucleus of a uranium atom with a mass number of 238 is 146, but the number of protons is only 92. As the mass number of the atomic nucleus increases, its instability gradually increases. At present, only about 105 relatively stable atoms are known, and the lifetime of the last atoms in the Periodic table is quite short. Atoms with the same number of protons but a different number of

neutrons are called **isotopes**. The stability of the isotopes is lower than that of the atoms themselves and decreases rapidly as the number of neutrons in the nucleus increases. Each atom has one to several relatively stable isotopes, which are mixed in very small fractions with the main atoms. This explains the non-numerical value of the atomic masses of the elements in the Periodic table.

In addition to protons and neutrons, atomic nuclei also include a large group of low-stability elementary particles, the so-called **mesons**, participating in one way or another in intranuclear interactions. Electrically, mesons can be either charged or neutral. The mass of mesons fluctuates within a wide range. At first it was thought to be between the mass of the electron, which is $9.1 \cdot 10^{-31}$ kg (0.51 MeV/s^2), and the mass of the proton. However, later it turned out that the mass of the overwhelming number of mesons is greater than the mass of the proton, exceeding it by several times. In some cases, it is even more than an order of magnitude. The spin of mesons is integer or zero, i.e. mesons, unlike nucleons, are bosons.

One of the most important properties of strong interactions is their **isotopic invariance**, that is, their independence from the charge state of nucleons. This means that both two protons and two neutrons, or a proton and a neutron, interact in the same way. Strong interactions are also characterized by a very small radius of action, equal to about 10^{-15} m. They result from the exchange of nucleons by **pi-mesons**, also called **pions**. Pions are part of the so-called meson "cloud" with which nucleons are always shrouded The existence of pions was postulated in order to explain the short-range effects of nuclear forces by the Japanese physicist Hideki Yukawa in 1935. Pions were experimentally detected in 1948 - 1950 at the particle accelerator in Berkeley (USA). Pions have spin zero and are a group of bosons consisting of three elementary particles, two electrically oppositely charged and one neutral. The mass of charged pions is 140 MeV/s^2, and the neutral one is 135 MeV/s^2.

Mesons and baryons belong to the class of **hadrons**. Hadrons are elementary particles composed of smaller particles, **quarks** and **antiquarks**. Quarks and antiquarks, in addition to the properties that characterize all elementary particles, also have specific properties that are described by special quantum numbers called **flavor** and **color**. In addition, quarks are also characterized by quantum numbers called **strangeness** (S), **charm** (C), and **beauty** (b). The concept of quarks was introduced in 1964 by American physicists George Zweig and Murray Gell-Mann. They assumed that all baryons are built from different combinations of the three fundamental quarks, u, d, and s (see Table 7.3), which differ from each other by their respective quantum numbers. Mesons consist of a quark and an antiquark.

Table 7.3.

Flavors of Quarks		u	d	s
Mass	MeV/s^2	4.1	5.12	8.7
	kg	$7{,}28 \cdot 10^{-30}$	$9{,}1 \cdot 10^{-28}$	$1{,}54 \cdot 10^{-29}$
Electrical charge in units of electron charge		+ 2/3	– 1/3	– 1/3
Isotopic spin	The quantity	1/2	0	1/2
	Projection	+ 1/2	0	– 1/2
Spin		1/2	1/2	1/2
Baryon charge in units of positron charge		1/3	1/3	1/3
Strangeness, S		0	0	– 1
Charm, C		0	0	0
Beauty, b		0	0	0

All hadrons are combined into one class and form a symmetry invariant with respect to the transformations of **internal** quantum numbers.

Quantum numbers, as mentioned above, are integers or fractions that define possible values of **physical quantities (states)** of quantum systems. The set of quantum numbers that comprehensively defines the

state of a quantum system is called **complete**. It was shown above (see section 7.4.1) that the state of an electron in an atom is determined by four quantum numbers, which are **eigenvalues of wave functions**. For example, the state of an electron in a hydrogen atom is determined by the principal quantum number, the orbital quantum number, the magnetic and spin quantum numbers. The same quantum numbers describe multi-electron atoms with a certain approximation. In addition to these quantum numbers, **internal quantum numbers** are also very important for description of quantum systems. The effect of these numbers is manifested in particle **interactions**. These include: conserved **electric** Q and **baryonic** charge B, conserved with some approximation in strong and electromagnetic interactions strangeness (S), charm (C), and beauty (b). Internal quantum numbers also include weakly conserved **isotopic spin** $\vec{I}$. In addition, elementary particles are also characterized by the already known **parity**, which remains unchanged in all interactions.

The isotopic spin characterizes the **isotopic symmetry** corresponding to the SU(2) group transformations and is part of the unitary symmetry corresponding to the SU(n) group transformations. Unitary symmetry is inherent to strong interactions and belongs to the group of **internal symmetries** of quantum systems. The value n, which characterizes the internal symmetry group, is the number of components that make up all possible combinations of the group. Each such group consists of similar particles with approximately the same masses and the same internal quantum numbers, including - spin, parity, baryonic charge, strangeness, charm, and beauty. Particles within each group with isotopic symmetry differ from each other by electric charge and mass. These groups are called **isotopic multiplets**. For example, nucleons form a **duplet**, pions form a **triplet**, etc. The **multiplicity** N is the number of members of the multiplet and is determined from the relation

$$N = 2\vec{I} + 1 \tag{7.140}$$

The value of isotopic spin is conserved for each group separately. So, for example, for nucleons $\vec{I} = 1/2$, since $N = 2$, and for pions $\vec{I} = 1$, since $N = 3$. Isotopic spin is a vector quantity. The projection of isotopic spin $\vec{I}_3$ is related to the electric charge of the particle Q by the Gell-Mann–Nishijima formula by the following relation

$$Q = I_3 + \frac{1}{2}Y, \tag{7.141}$$

where Y is the hypercharge of the particle.

$$Y = B + S + C - b \tag{7.142}$$

Thus, if the isotopic spin vector is unchanged, its projection has different values for each of the particles in the group. For example, for nucleons the projection has two values of +1/2 and -1/2, and for pions three values of +1, -1, and 0. Since the strong interaction does not depend on the value of the electric charge, it can be formally interpreted as invariance with respect to the rotation of the isotopic spin vector in conditional isotopic (charge) space or as the existence of the symmetry group SU(2). This group is defined by two varieties of quarks out of two.

Unitary symmetry, in turn, unites groups of isotopic symmetry, (in our case different groups of hadrons), into one class. It reflects an approximate symmetry with respect to changes in isotopic spin, strangeness, and mass, and is described by the symmetry group SU(3), since it is formed by the three above-mentioned varieties of quarks. Despite the approximate nature of this symmetry and the great difference between hadrons, especially fermions and bosons, it nevertheless allows us to combine nucleons and mesons into one class due to the symmetry of their composition. Based on SU(2) symmetry and the fact that the spin of all quarks without exception is 1/2, the following quark model of hadrons was created:

protons - *udd*

neutrons - *udd*

π^+ mesons - ud^*

π^-mesons - $u^* d$

etc. (the letter with the asterisk at the top indicates the antiparticle).

The whole variety of hadrons arises from combinations of different quarks and antiquarks that form bound states. In this case hadrons, which are formed from 3 quarks, have a total half-integer spin (1/2 or 3/2). These hadrons belong to the fermions and are part of the group of baryons or antibaryons. As mentioned above, all baryons without exception are assigned a baryonic charge equal to + 1, and antibaryons are assigned with a charge equal to – (– 1), which characterizes their interaction with each other. Hadrons consisting of 1 quark and 1 antiquark have a total spin equal to zero or 1. They belong, respectively, to bosons and form a subclass of mesons. Due to the fact that mesons do not interact with each other, but are only carriers of interactions, their baryon charge is 0. The composition of hadrons also explains the increased lifetime of baryons, whose structural elements are bound by strong interactions, and the short lifetime of mesons, in which the action of particles is compensated by the counteraction of their antiparticles.

The electric charge of hadrons is known to be 0 or ±1. This means that the charge of the quarks is half-integer, with the quarks of u type assigned a charge of + 2/3, and quarks of *d* and *s* type - a charge of – (– 1/3). With this in mind

$$e_p = + 2 / 3 + 2 / 3 – 1 / 3 = + 1$$

$$e_n = + 2 / 3 – 1 / 3 – 1 / 3 = + 0$$

$$e_{\pi+} = – 2 / 3 – 1 / 3 = – 1$$

$$e_{\pi–} = – 2 / 3 – 1 / 3 = – 1$$

The SU(3)-symmetry is based on the independence of the interaction energy from the type of quarks that form hadrons.

All 3 quarks are characterized by the same spin equal to 1/2, the same values of baryonic charge B = 1/3, charm C = 0, and beauty b = 0.

Baryons, as we have seen, consist of three quarks. Often two of them, and sometimes all three, are the same. In the case of two identical quarks, their spins are antiparallel, and the Pauli exclusion principle is not violated. However, hadrons of three identical quarks could not be realized, although in practice they exist. This led to the introduction of another characteristic of quarks called **color**. The quarks, which are identical in flavor, differ from one another in their color. Such a strange name for this characteristic arose due to the fact that it is always absent in hadrons. In other words, the quarks, when united, as it were, compensate for their color, or, as they say, decolorize each other. The mesons are simpler. They always consist of a pair - a quark and an antiquark. Therefore, if one of them has a color, the other has an anti-color, as a result the colors are mutually cancelled out. In baryons, not two, but three quarks must be discolored. But it is the colors that have a similar property. For example, red, green, and purple combine to give a discolored white. Hence, by analogy, the name of this characteristic arose. The notion of color also helped to explain the well-known fact that quarks are undetectable in the free state. The fact is that the interaction of quarks is carried out by their exchange of gluons, which, by analogy with the quanta of the electromagnetic field, photons, are massless particles with spin equal to 1. In this case the source of the field of strong interactions by the same analogy is the color charge. However, the specifics of the field of strong interactions, unlike the electromagnetic field, is that the charge, in this case the color charge, is attributed not only to interacting particles, fermions, in this case quarks, but also to particles carrying the interaction, bosons, in this case gluons. This means that for small distances, the effective color charge of quarks is compensated by the color charge of gluons and, accordingly, decreases with decreasing

distance. Therefore, at small distances quarks behave as quasi-free particles. On the contrary, when two quarks are mutually removed, their color charge and interaction energy increase so much that it leads to the birth of a quark-antiquark pair, which decolorizes the separated quarks and combines them into a colorless hadron. It follows that obtaining quarks in the free state, at least at currently available and predictably distant future energies, is in principle impossible.

Hadrons consisting of u, d, s quarks whose charms are zero cannot be charmed. Therefore, the discovery of charmed mesons, mesons with hidden charms, and charmed baryons led to the need to expand the family of quarks, and to introduce the so-called charmed ($c \neq 0$) and beautiful ($b \neq 0$) quarks. At the same time, the existence of new families of hadrons with SU(4) and SU(5)-symmetry was postulated. A four-quark hadron was recently discovered at the Large Hadron Collider.

One of the most common structural models of the atomic nucleus is the shell model. The nucleon of a nucleus, like the electrons of an atom, are quantum particles. Each of them is in a particular quantum state characterized by an energy level, i.e., the principal quantum number n, the orbital moment J and its projection m on one of the coordinate axes, spin, parity, etc. According to Pauli's exclusion principle, each energy level can contain 2 $(2J + 1)$ identical nucleons, forming a nucleon shell, so that a completely filled shell would contain 2 n^2 nucleons. However, unlike the shells of the atom, which consist of the same particles, electrons, nucleonic shells are formed by different particles and therefore we distinguish between proton and neutron shells separately. As a result, the laws of shell filling by nucleons, which are present for atoms, are violated.

The number of protons and neutrons in the nuclei is determined by a series of so-called **magic numbers**. For neutrons they are 2, 8, 20, 28, 40, 50, 82, 126, and for protons they are 2, 8, 20, 28, 50, 82. Nuclei with both the number of protons and the number of neutrons as magic

numbers are particularly stable and strong, reminiscent of the chemical stability of noble gas atoms. The magic numbers observed in the experiment correspond to quantum states of quasiparticles moving in a rectangular potential well with spin-orbit interaction.

In a completely filled shell, the moments of nucleons are mutually compensated, and the total momentum of the shell, as well as their spin, is zero. In each case, when the number of neutrons or protons in the shell becomes equal to the number of completely filled shells, there is a jump change in some characteristics of the nucleus (e.g., bonding energy), which is equivalent to the law of periodicity of changes in the properties of atoms. This is also indicated by many of the magic numbers, such as 2; 8; 20(18+2); 28(20+8); 82(32+32+18), etc. The shell model also made it possible to explain the spins and magnetic moments of nuclei.

It should be emphasized, however, that the shell structure of nuclei is much weaker than that of atoms. This can be explained, first, by the absence of a cementing central force in the nucleus, which results in increased mobility and mutual collisions of nucleons, and, second, by the fact that the motion of nucleons can be discussed with even greater convention than the motion of electrons.

The nucleons in the nucleus create a collective formation in which the individual properties of each individual nucleon are so weak that they are rather contingent. This made it possible, by analogy with the crystal structure of metals, to complement the shell model with the so-called generalized model, according to which the nucleus has a sufficiently stable framework consisting of filled shells. The remaining outer nucleons that did not enter the framework form a Fermi liquid, which in the ground state is treated as an ideal Fermi gas. In the excited state, one or two quasiparticles move to higher energy levels. By releasing orbits inside the fermionic sphere, they can interact both with each other and with the resulting hole in the lower shell. The hole, in turn, can move, going from one state to another. The difficulty of the theory is that the

interaction of quasiparticles and holes is actually quite large, which contradicts the initial statements and can lead to inaccurate final results. To increase the accuracy of the theory, the method of successive approximations is used, which consists in considering the motion of quasiparticles only as the first approximation, and the subsequent approximations are obtained using empirically chosen model parameters.

7.4.10.1. Intranuclear processes.

One of the important characteristics of a nucleus is its mass, which is determined by the mass of its constituent nucleons. However, the total mass of nucleons bound in the nucleus is always less than their total mass in the free state. This is not surprising, since, according to the definition, mass is a quantity proportional to the lag time of transmission of interactions. In the case of a bound state of particles, the value of this time is always less than in their free state. The difference of these masses Δm is called a **mass defect**. It follows that the absolute value of Δmc^2 determines the **binding energy**, i.e., the amount of work that must be done to transfer nucleons from a bound state to a free state. Since the total mass of the decaying particles increases, their equivalent energy also increases. In other words, the decay process is accompanied by energy absorption. On the contrary, energy is released during fusion, i.e., the joining of free nucleons into a nucleus.

From this point of view, we can conventionally assume that all types of energy are divided into positive energy of relative motion of structural elements of loosely connected systems, which is a measure of motion of these elements, and negative energy of connection of elements connected into a system, which determines, on the contrary, the measure of their relative immobility. In this sense, gravitational energy or mechanical potential energy is bonding energy, and kinetic energy or internal energy of motion of particles is energy of motion. If we denote the binding energy $\mathcal{E}_b$, the masses of the proton, neutron, and nucleus, respectively m_p, m_n, m_N, then

$$\mathcal{E}_b = [Zm_p + (A - Z)\, m_n - m_N]\, c^2.$$

The amount of binding energy per nucleon is called the **specific binding energy**. Figure 7.28 shows a graph of the dependence of the specific binding energy on the mass number (the number of nucleons in the atom). The graph shows that for the vast majority of nuclei, the specific binding energy is between 8 and 9 MeV/nucleon.

Only for a small number of light nuclei and nuclei with a large mass number is it significantly smaller. The nuclei of the elements at the end of the Periodic table are less stable, and very heavy nuclei are unstable.

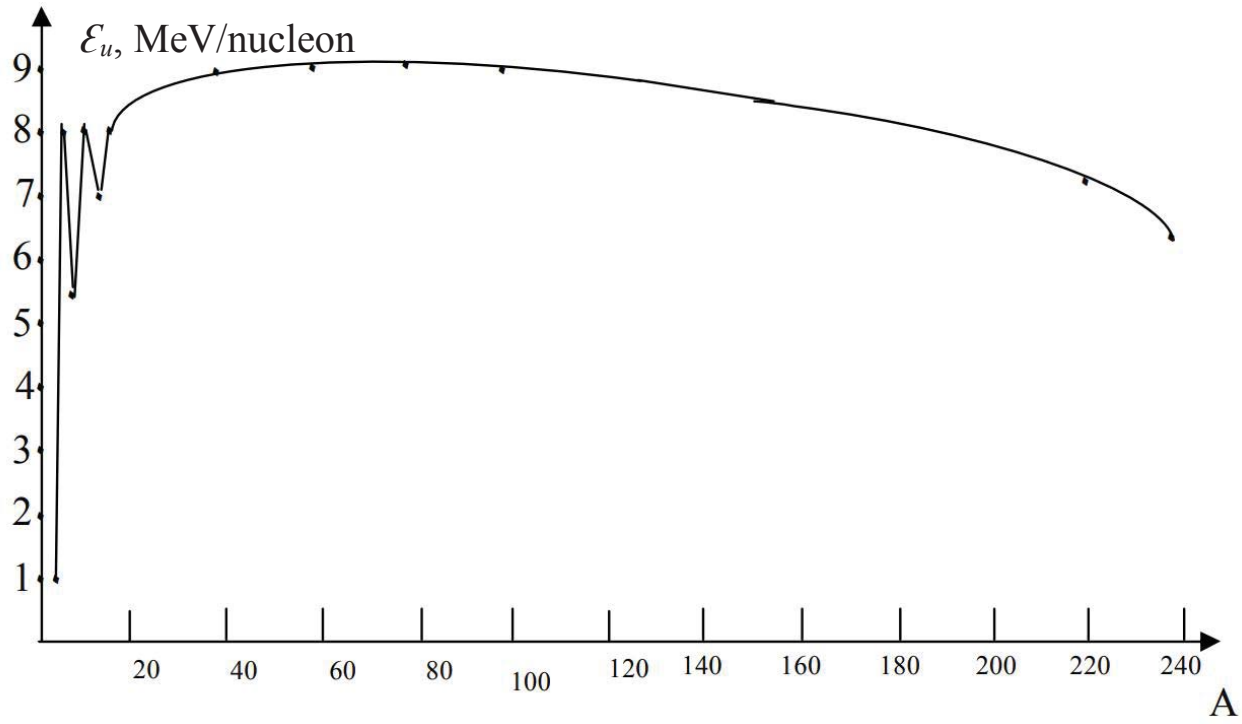


Figure 7.28.

There are two possible ways of releasing energy (obtaining nuclear energy) during intranuclear processes: the **fusion of light nuclei** that we discussed above, and, on the contrary, the **fission of heavy nuclei**. In both cases, nuclei with higher specific bonding energies and,

consequently, lower nucleus energies are formed, which entails a mass defect and excess nuclear energy.

Low-stable isotopes of stable nuclei, as well as heavy, low-stable nuclei, are subject to spontaneous **radioactive decay**, in which the nuclei of some chemical elements, called the **parent radionuclide**, are transformed into the nuclei of other elements, called the **daughter nuclide**. The process of radioactive decay is called **radioactivity** and is usually accompanied by the emission of particles and quanta of electromagnetic energy. A distinction is made between **natural radioactivity**, which occurs in nature without human intervention, and **artificial radioactivity**, produced by bombardment of nuclei with energetic particles. In the latter case, **nuclear processe**s are accompanied by **nuclear decay**. The natural radioactivity of nuclei was discovered by the French physicist Henri Becquerel in 1896, and the artificial one was discovered by the Joliot-Curie couple in 1936.

Let us denote by N the number of nuclei of a given type at time t, and by dN the decrease in the number of spontaneously decaying nuclei over time dt. It is quite obvious that

$$-dN = \lambda N dt,$$

where λ is the radioactive decay constant.

By integrating the last expression, we obtain

$$N = N_0 e^{-\lambda t}, \qquad (7.143)$$

where N_0 is the initial number of nuclei;

The inverse of the radioactive decay constant, $\tau = 1/\lambda$, is called the **lifetime** of radioactive nuclei.

The intensity of the radioactive decay process is measured by a value called the **half-life** $T_{1/2}$. The half-life is the **time** during which **half of the initial number of nuclei** decays. According to this definition

$$\frac{N_0}{2} = N_0 e^{-\lambda T_{1/2}};\ \ T_{1/2} = \frac{\ln 2}{\lambda};\ \ \tau = \frac{T_{1/2}}{\ln 2}.$$

The half-lives of various radioactive elements vary widely, from 10^{-7} s to 10^{10} years or more.

Depending on which particles or quanta of energy are emitted in the radioactive decay of nuclei, a distinction is made between α-decay, which emits nuclei ^4_2He called α-particles, β-decay, which emits electrons called β-particles, and γ-decay, which emits hard γ-rays, i.e. high energy photons.

To write the equations of nuclear processes (reactions) we denote nuclei by ^A_zX. Let us also denote α-particles by the symbol for helium nuclei, β-particles by the symbol for electrons. Finally, we denote neutrons by n, protons by p, neutrinos by ν, antineutrinos by ν*. On the bottom left of the notation of charged particles we will indicate the value and sign of their charge, and for neutral particles we will indicate 0. On the top left of the particle notation we will indicate its mass number, and for leptons and antileptons (electrons, neutrinos, etc.) we will indicate 0. Let us also call the original (decaying) nucleus the **parent** nucleus, and the nuclei obtained during the decay of the mother nucleus the **daughter** nuclei. Nuclear reactions occur in this case according to the **displacement rule**, according to which the value of the total charge Z and the mass number A of the daughter nuclei must equal the value of these values of the parent nucleus.

During α-decay, according to the rule of displacement, the charge of the daughter nucleus decreases by 2 units (the charge of the helium nucleus), and the mass number decreases by 4 units. The nuclear reaction of decay, e.g. of a uranium nucleus, looks like

$$^{238}_{92}U \rightarrow\ ^4_2He +\ ^{234}_{90}Th, \qquad (7.144)$$

In some cases, alpha decay is also accompanied by the emission of a γ-quantum energy. In all cases, the emission of alpha particles is associated with the tunnel effect.

The nuclear reaction in the β-decay of, for example, radioactive cobalt, by analogy and taking into account the displacement rule, can be written as follows

$$^{60}_{27}Co \rightarrow \ ^{60}_{28}BNi + \ ^{0}_{-1}e + \ ^{0}_{0}\upsilon^{*} + \gamma, \qquad (7.145)$$

As can be seen from the above reaction, in β-decay the nucleus emits an electron and antineutrinos, which are not contained in it, but are born during the reaction itself. As will be shown below, this decay is a consequence of the weak interaction, in which the neutron transforms into a proton with the emission of an electron and an antineutrino. The result of this is the emergence of the nickel atom, whose daughter nucleus contains one more proton and one less neutron than the parent nucleus of the cobalt atom, while maintaining the mass number. Unlike α-decay, which has a discrete energy spectrum, in β-decay it is continuous. In positron decay (we recall that the positron is an antielectron), the mother nucleus transforms into the daughter nucleus, emitting a positron and a neutrino. In the process of positron decay, a positron and neutrino are produced, which is equivalent to the transformation of a proton into a neutron, as opposed to electron decay. Another type of β-decay is associated with the capture by the nucleus of an electron from the nearest K-shell of the atom (K-capture). When united with one of the protons of the nucleus, the captured electron transforms it into a neutron and emits a neutrino.

It is interesting to note that it is the β-decay that is associated with the history of neutrino discovery. The point is that, as the measurements showed, the total energy of the original particles was always greater than the energy of the particles produced in the decay process. In other words, beta decay violated the law of conservation of energy, some of the energy disappeared somewhere silently. Analyzing the situation, Pauli

hypothesized in 1932 that beta decay gives birth to another neutral very small particle, the neutrino, with a spin of 1/2 and a rest mass equal to zero, which does not remain in the nucleus, but is carried away from it at the speed of light. Due to the enormous penetrating power of neutrinos, it was not detected for a long time. Only in 1956, its existence was confirmed experimentally by American physicists Frederick Reines and Clyde Cowan. Research by Soviet physicists in the 1980s, confirmed by a number of American physicists, indicates that the neutrino rest mass is not zero and is in the range of 10-30 MeV/s^2. However, these data are not yet considered completely reliable.

Gamma radiation arising from the radioactive decay of atomic nuclei is in the short-wave part of the spectrum and is characterized by a wavelength of 10^{-10} m. The passage of γ-rays through matter is accompanied by an internal photoelectric effect, Compton scattering, and photoproduction of particle-antiparticle pairs. The photoelectric effect arises as a consequence of the interaction of a γ-quantum with the electrons of the atom's inner shells, so it results in X-rays. Compton scattering appears only at gamma-quantum energies of the order of 2 MeV and leads to the absorption of gamma rays by matter. At energies higher than 1.1 MeV, an electron-positron pair birth is also possible.

The effects of radiation on matter, including living organisms, are characterized by the dose of radiation absorbed and the permissible dose at which irreversible destruction of the substance or organism does not yet occur. The **absorbed dose** is the amount of absorbed energy per unit mass of the irradiated substance. The unit of absorbed dose is 1 gray (Gy) = 1J/kg. The smaller unit is 1 rad = 10^{-2} Gy. Direct measurement of the absorbed radiation dose is not possible, so we measure the associated **amount of ionized substance charges** that arise under its exposure, which is called the **exposure dose of radiation** and is measured in Cl/kg. The unit of exposure radiation is 1 Cl/kg or 1 roentgen (1 R), equal to $2.58 \cdot 10^{-4}$ Cl/kg, which corresponds to the formation of $2.08 \cdot 10^3$ ion pairs/m^3. The absorbed dose of radiation produces different biological

effects depending on the type of radiation, so the value of the coefficient of relative biological effectiveness (RBE) in relation to the effect of γ-radiation is also introduced. The unit of RBE is 1 roentgen equivalent man (rem) = 10^{-2} J/kg and is equal to a radiation dose that produces the same effect as 1 R of gamma radiation.

The energy spectrum of γ-radiation is discrete and, according to the uncertainty relations, its levels are broadened by the value of $\Delta\mathcal{E} = \frac{\hbar}{\tau}$, where τ is the duration of a nucleus in a state with a given energy. As in the case of the atom, the level of the unexcited, ground state does not broaden. The uncertainty of the energy of excited states leads to non-monochromaticity of the radiation, which is called the **natural width of the emission line**.

In 1958, the German physicist Rudolf Mössbauer established that it is possible for atomic nuclei bound in a solid to absorb or emit γ-rays without changing the internal energy of the body. This is explained as follows. When a nucleus emits or absorbs a γ-quantum, the system containing the nucleus acquires a recoil pulse, with the emission and absorption lines shifting in different directions from the main line corresponding to the given quantum transition and broadening. The recoil energy turns into the energy of thermal fluctuations of the crystal lattice, which are equivalent to the set of additional phonons born in this process. When the recoil energy per nucleus is less than the average energy per phonon of a given crystal, not every act of absorbing a γ-quantum results in a phonon birth or, similarly, in a change in the internal energy of the body. This phenomenon is called the **Mössbauer effect** or nuclear resonance. The resonance absorption of gamma rays is strongly expressed if the energy of oscillatory motion of the crystal lattice is not superimposed on it, and the recoil energy is transferred not to a single nucleus, but to the lattice as a whole. The broadening of the emission lines is natural in this case and is exceptionally small, on the order of 10^{-10} eV. The probability of nuclear resonance increases as the energy of the γ-

transition decreases, as the temperature of the crystal decreases. Therefore, to observe it, the temperature of the crystal is reduced to the temperature of liquid helium or liquid nitrogen. Due to its high sensitivity, the Mössbauer effect has been widely used to measure small energy shifts of γ-rays caused by various influences on the emitting or absorbing nucleus.

7.4.10.2. Nuclear reactions.

Nuclear reactions differ in:

- the type of particles involved - neutrons, charged particles (protons, electrons, deuterons, tritons, α-particles), γ-quanta, π-mesons, etc;

- the nature of the nuclei causing the reaction, i.e., reactions occurring on light nuclei with a mass number A up to 50, on medium nuclei when $50<A<100$, or on heavy nuclei when $A>100$;

- particle energies; at low energies, on the order of 1 eV, medium energies, up to 100 MeV, high energies, over hundreds and thousands of MeV;

- the nature of nuclear transformations - with emission of neutrons, with emission of charged particles, with emission of γ-quanta, with scattering of particles, nuclear fission and fusion.

Reactions induced by neutrons are of particular importance. Unlike charged particles, neutrons do not experience Coulomb repulsion from the nuclei they bombard, so they easily penetrate nuclei and are most effective in terms of the nuclear reactions they cause. A distinction is made between fast neutron reactions, which have a de Broglie wavelength shorter than the nucleus radius and an energy of 0.1-50 MeV, and slow neutron reactions, which have an energy of no more than 100 keV. Thermal neutrons with energies from 0.005 to 0.5 eV, cold neutrons

with energies 10^{-3} to 10^{-4} eV, and ultracold neutrons with energies 10^{-7} to 10^{-8} eV are also distinguished.

When fast neutrons collide with a nucleus, there is a scattering-type reaction, where the absorbed neutron corresponds to the emitted neutrons, or a reaction with the formation of charged particles. The return of an excited nucleus to the ground state is accompanied, as a rule, by γ-radiation.

Nuclear reactions under the action of slow neutrons are the most effective, and from the point of view of nuclear energy generation, the **fission reactions** of heavy nuclei are of primary interest.

The fact is that the daughter nuclei arising in the process of fission are the nuclei of the middle elements of the Periodic Table. Each contains a relatively smaller number of excess neutrons than the original heavy parent nucleus. As a result, the fission products may contain more than one neutron per decayed original atom. Each of these neutrons, in turn, causes nuclei fission so that there is more than one primary neutron per atom fission, etc.

The process of fission under the influence of neutrons occurs as follows. Let a heavy nucleus capture a neutron and get excited. As a result of excitation, the nucleus can emit a γ-quantum and return to the initial, ground state. However, much more often it deforms and, under the influence of the Coulomb proton repulsion forces, disintegrates into two fragments of unequal masses, which, as already mentioned, are saturated with excess neutrons and are, therefore, unstable. The fragmented nuclei thus excited emit excess neutrons, which are called secondary neutrons, and go to a steady state.

The number of secondary neutrons per fission is different for different nuclei. For the ^{235}U nucleus, for example, it is quite large and averages 2.5 to 3. At the same time, thermal nuclear energy Q is released, which is calculated by the following formula

$$Q = \left[(M_X + m_x) - \left(M_{Y_1} + M_{Y_2} + m_y\right)\right]c^2, \quad (7.146)$$

where M_X, M_{Y1}, M_{Y2} are the masses of the parent and daughter nuclei;

m_x, m_y are the masses of primary and secondary particles.

For the ^{235}U nucleus, which decays into ^{140}Ce and ^{94}Zr nuclei with one primary and two secondary neutrons, Q = 208 MeV. (For comparison, chemical combustion reactions emit only 4 MeV, and α-decay emits 5 MeV.)

Fission reactions also benefit from the fact that they can be chain reactions, i.e., self-sustaining. If one primary neutron gives rise to at least two secondary neutrons, which in turn give rise to four tertiary neutrons, etc., the result is a spontaneous avalanche-like increase in the number of fissile nuclei, and the reaction proceeds with an instantaneous increase in intensity. It should be kept in mind, however, that not all neutrons cause fission, as some of them leave the core, where fission occurs, before they are captured by fissile nuclei, or get into the fissile atoms of the core's structural materials. In addition, some neutrons cause radiation capture instead of fission, etc. In this regard, an important characteristic of a chain reaction is the so-called neutron **multiplication factor** K, which is defined as the ratio of the number of neutrons in a given generation to its number in the previous generation. Let in a given generation the number of neutrons is N, then in the next generation it is KN. Therefore,

$$dN = \frac{N(K-1)}{\Delta t}\,dt.$$

By integrating, we get

$$N = N_0 e^{\frac{(K-1)t}{\Delta t}}, \quad (7.147)$$

where Δt is the average lifetime of one neutron generation.

If K < 1, then N < N0 , and division stops with time. This corresponds to the attenuating mode of the reaction. At K > 1, the reaction

builds up avalanche-like, and energy is released in the form of an **explosion**. This mode is implemented in the atomic bomb. Finally, when K = 1, the reaction is stably self-sustaining. This mode is called **critical** and is implemented in so-called nuclear reactors. Critical mode characteristics (size, mass, etc.) are also called critical characteristics.

A distinction is made between **controlled** and **uncontrolled** chain reactions. In the uncontrolled reaction of an **atomic bomb** (warhead) its active volume is divided into two parts, which are mutually separated.

The mass of each part must be lower than the critical mass, i.e. correspond to the attenuating, so-called **subcritical mode**. The weight of the parts is chosen so that when they are combined, the total mass is greater than the critical mass. Primary neutrons are created by the natural radioactivity of nuclear fuel (fissile material), which is ^{235}U or ^{239}Pu (the process of obtaining naturally inexistent ^{235}U from naturally occurring ^{238}U is called **uranium enrichment**).

Devices with a controlled chain reaction designed to produce atomic energy used for industrial and domestic purposes are called **nuclear reactors**. A nuclear reactor consists of a core containing nuclear fuel (fissile material), a moderator, a reflector, a cooling system, a control and protection system, and a control panel. As nuclear fuel, reactors use pure and highly or not highly enriched nuclides (heavy nuclei) ^{235}U, ^{239}Pu, ^{233}U. The core is usually made as a set of fuel elements (fuel rods) in which nuclear fuel is placed. This is where nuclear fission takes place. The heat that is released during the fission reaction is removed from the fuel rods by washing them with heat transfer medium. Heat transfer media can be water, heavy water, gas, liquid metal and also organic substances. Helium or carbon dioxide is used as gas heat transfer medium. Sodium, alloys of lead with bismuth or sodium are used as the liquid-metal ones. Gasoil, biphenyl, biphenyl mixtures, etc. are organic media. Circulating in a special circuit placed in the core, the heat transfer medium is heated and transmits heat to the working body, most often

water. In so-called **homogeneous** reactors, the heat transfer medium is a homogeneous mixture of nuclear fuel and moderator in the form of a suspension or liquid solution. In **heterogeneous** reactors, the nuclear fuel is in the form of rods, tubes, blocks, etc., which are not in contact with the moderator and heat transfer medium and are distributed among reactor parts made of other substances. The core is surrounded by a neutron reflector designed to reduce neutron leakage and equalize heat release over the volume of the core. The reflector materials used are the same as those used for heat transfer media or special substances. The moderator is designed to convert the fast neutrons born in the fission process into slow or thermal neutrons. Graphite, heavy water (D2O), beryllium, or its oxides are used as moderators. The reactor is controlled by moving rods of cadmium or boron, which are strong neutron absorbers. The deeper the rods are pushed into the core, the lower the intensity of the nuclear reaction and the reactor power. The movement of the rods is controlled automatically. In the event of an emergency, the rods automatically retract fully and the nuclear reaction stops.

The process of nuclear fission is accompanied by the production of radioactive daughter nuclei, the emission of energetic neutrons, as well as α, β and γ radiation. To protect people and equipment from radioactive radiation, the reactor is submerged in water to a depth of about 1 m and encased in a thick concrete casing.

Both fast- and slow-neutron reactors are currently in use.

The synthesis of light nuclei, as was shown above, is also accompanied by the release of energy, which, moreover, exceeds the corresponding fission energy per nucleon by more than 4 times. It is also essential that there may not be any neutron emission and associated radioactivity during fusion. At the same time, nuclear fusion also poses a serious problem related to the need to overcome the Coulomb repulsion of protons. For example, in order to unite two deuterium nuclei, they must be brought closer together to a distance of $2 \cdot 10^{-15}$ m. The potential

energy of Coulomb repulsion at this distance is 10 keV. This corresponds to a temperature of about $2 \cdot 10^8$ K. This reactions are called **fusion reactions**. The first uncontrolled fusion reaction of deuterium and tritium nuclei into helium was carried out in the form of a hydrogen bomb explosion in the Soviet Union in 1953. It used a synthesis reaction according to the following scheme

$$D_1^2 + T_1^3 \rightarrow He_2^4 + n_0^1 + 17{,}6\,MeV \qquad (7.148)$$

The temperature necessary for the fusion reaction to take place was created by the preliminary explosion of a small atomic bomb, thus serving as the detonator of the hydrogen bomb. The fusion energy heated the plasma to a very high temperature, at which the fission chain reaction of the bomb shell, made of low-deficit ^{238}U, began. The result was an explosion that was many times more powerful than an atomic bomb. This explosion was also accompanied by intense radioactive emissions. The source of tritium, which, because of its radioactivity, is practically absent in nature, was the reaction of its reproduction from the lithium deuteron when irradiated with ^{6}Li neutrons according to the following scheme

$$^6_3Li + n = {}^3_1T + {}^3_2He$$

The thermonuclear fusion reaction of hydrogen nuclei is a source of solar energy. This reaction is very slow with little heat release. This disadvantage, which makes this reaction inapplicable in terrestrial conditions, is compensated for in the Sun by the huge size of the solar reactor.

The high temperature of the fusion reaction is one of the reasons why it has not yet been possible to realize a controlled fusion reaction on an industrial scale. The problem is not only to create, under earthly conditions, such a high temperature at which all substances turn into plasma, but also to keep this plasma in the vessel in a stationary state. First, the contact of the high-temperature plasma with the walls of the vessel leads to its almost instantaneous cooling, and second, it just as

quickly destroys the vessel, turning it into plasma. In addition, a fusion controlled reaction in general only makes sense if it is energetically beneficial, that is, if the resulting fusion energy is greater than the energy expended to excite the high temperature and contain the plasma. The conditions necessary for the realization of a regulated **self-sustaining** thermonuclear reaction are called the **Lawson criterion**. The characteristics of this criterion are the **plasma confinement** parameter $n\tau$, (where n is the concentration of particles in the plasma during its confinement time τ) and the plasma temperature T. Calculations have shown that for the passage of a controlled fusion reaction at a temperature $T = 2 \cdot 10^8$ K and the synthesis of a mixture of deuterium - tritium is provided under the condition that $n\tau > 10^{14}$ c/m^3. Tritium is produced using a small atomic reactor.

The lower limit of the plasma dispersal time in the absence of plasma confinement devices and a concentration comparable to the concentration of atoms in the volume of a solid body allows a nuclear fuel fusion reactor to be made in the form of small grains. Laser pulses of about 50 GW, joule heat of electric current, heat generated by electrodynamic plasma compression, high-frequency heating, bombardment of nuclei with fast particle beams, or electron beams can be used to heat the nuclear fuel to the desired temperature.

Keeping plasma inside the vessel, according to the idea of Soviet physicists Igor Tamm and Andrei Sakharov, is done by isolating it relative to the walls of the vessel (pushing it away from the walls by a powerful magnetic field). This idea was put into practice in fusion reactors such as **Tokamak**, **Torsatron**, and **Stellarator**. The name of the reactors is related to the way the plasma is heated and the insulating magnetic field (toroid, coil, magnet) is created, as well as its stabilization. Plasma in these reactors is created inside a toroidal coil or chamber by an electric current flowing through the coil, and the magnetic field by a magnet or two toroidal current windings (the field of one pulls the plasma

away from the toroidal chamber walls, while the other, helical, stabilizes the plasma cord).

The biggest disadvantage of first-generation reactors is their radioactivity. More promising are reactors that use mixtures of deuterium and helium three or hydrogen-boron as a fuel. In these reactors, the fusion reaction is neutron-free and their radioactivity is close to zero.

7.4.11. Elementary particles

Elementary particles include structural elements of physical fields and atoms of physical bodies, as well as some particles that fill interatomic, intermolecular, and interstellar spaces. They were called elementary because at first they were considered to be structureless point formations (see Sections 1.2.1., 1.2.2.). Although, from a modern point of view, real objects can be neither structureless nor pointless, their name has been preserved.

The number of varieties of elementary particles discovered to date is more than 350, and with the increasing power of the particle accelerators with which they are studied, more and more particles are discovered. However, all elementary particles differ little from each other and are combined into a small number of groups of similar particles.

As it was mentioned above (see section 7.4.2.1.) elementary particles depending on the parity of the wave function describing them are divided into particles (antiparticles) of matter (fermions) with odd wave function and half-integer spin and into field particles (bosons) with even wave function and even as well as zero spin (see also section 7.5.). Fermions, unlike bosons, interact with each other, with their interactions carried by bosons. Depending on which fundamental interactions they participate in, all fermions are divided into two groups of similar particles - quarks (antiquarks) and leptons (antileptons). Quarks participate in all

fundamental interactions, leptons do not participate in strong interactions, and neutrinos are also not involved in electromagnetic interactions.

There are three families (generations) of fermions (antifermions). Each family consists of two leptons (antileptons) - (electron, electron neutrino), (muon, muon neutrino), (tau lepton, tau neutrino) and two quarks (antiquarks) - (u,d), (c,s), (t,b). In the conventional notation for elementary particles, these families are given below.

1st family	2nd family	3rd family
e^{-} (e^{+})	μ^{-} (μ^{+})	τ^{-} (τ^{+})
υ_e $(\upsilon_e{}^{*})$	υ_μ $(\upsilon_\mu{}^{*})$	υ_τ $(\upsilon_\tau{}^{*})$
u (u^{*})	c (c^{*})	t (t^{*})
d (d^{*})	s (s^{*})	b (b^{*})

The muon and tau-lepton of the second and third families are extremely similar in their properties to the electron, but differ from it in their much larger mass and very low stability. On the contrary, the muon and tau neutrinos, while also maintaining high stability, like the electron neutrino, have much higher mass. The same can be said about quarks of different families. It is also characteristic that as the family number increases, the masses of similar particles increase. The simplest explanation for the existence of different families of fermions comes down to the assumption that quarks and leptons are composite particles and all subsequent families (generations) are excited states of the first family. Although the structural nature of elementary particles is not in doubt, the composite nature of leptons and quarks remains highly questionable and unconfirmed.

The entire substantive atomic Universe is formed by particles of the first family. In particular, protons and neutrons, which form the nuclei of all atoms, are, as mentioned above, hadrons composed of different combinations of 3 (u,d)-quarks, and pi- mesons, which participate in their interaction, are also composed of two (quark-antiquark) (u,d)-quarks.

Electrons are part of all atoms and, in addition, together with the electron neutrino, participate in all weak interactions.

The role in the Universe of particles of the second and third families, as well as their compounds (baryons and mesons, which include quarks and antiquarks from the second and third families), has not yet been established.

Four kinds of bosons correspond to the four fundamental interactions - gluons (particles carrying strong interactions), photons (carrying electromagnetic interactions), vector bosons (carrying weak interactions), and hypothetical gravitons (carrying gravitational interactions). In addition, pi-mesons, which are bosons with zero spin, also take part in the transfer of strong interactions. It is possible that mesons result from the emission of baryons during their transitions from excited states, similar to the way photons are emitted by electrons during similar transitions in the atom.

Elementary particles, due to their extremely small size (less than 10^{-15} m) cannot be directly observed. However, very reliable methods have now been developed for their indirect observation and study of their properties. As early as 1912, the English physicist Charles Thomson Rees Wilson invented a device for observing not the particles themselves, but the traces of charged particles (tracks), which was called the **Wilson cloud chamber** after his name, or just a cloud chamber. The principle of operation of the Wilson cloud chamber is based on the phenomenon of condensation of supersaturated steam (see Section 2.7.3.3) on (ions), and the formation of liquid droplets along the trajectory of the charged particle movement. Particle tracks are photographed by several cameras to obtain a stereoscopic (three-dimensional) image. The cloud chamber is placed in a magnetic field, which, depending on the charge and momentum of the particle, curves its trajectory, and, consequently, the track, accordingly. The nature of a particle and its properties are determined by the magnitude of its path, the radius of curvature of the

trajectory, momentum, etc. The momentum P is determined by the formula

$$P \cdot c = \frac{Hr}{\cos\varphi},$$

where H is the magnetic field strength;

r is the radius of curvature of the trajectory;

φ is the angle between the field and momentum directions

c is the speed of light in a vacuum;

The Wilson cloud chamber was replaced in the 1950s by the more advanced bubble chamber. The principle of operation of this chamber is based on boiling of superheated liquid (see item 2.7.3.3.) near the particle trajectory. When a charged particle passes through a superheated liquid, a track appears along its trajectory, consisting of nucleated boiling centers and bubbles formed on them, reaching the size of 30-300 μm. Tracks are photographed when they are illuminated by a pulsed light.

The advantage of the bubble chamber over the Wilson chamber is that it allows one to register the acts of interaction of the particles under study with the nuclei of the working liquid, as well as the acts of decay of the particles in weak interactions. Liquid hydrogen or deuterium, mixtures of neon and hydrogen, propane, freon, xenon mixed with propane, etc. are used as the working fluid. Atomic nuclei of the working fluid play the role of a target. The properties of particles are determined by the magnitude of curvature of their trajectory in the magnetic field, the thickness and length of track lines, etc.

Charged particle accelerators are a powerful means of studying, producing, and detecting new elementary particles. American and Soviet physicists (Ernest Lawrence, Donald William Kerst, Vladimir Veksler, Edwin Mattison McMillan, etc.) participated greatly in the development

of such particle accelerators. In particle accelerators, charged particles are accelerated using an electric or electromagnetic field of high frequency to very high speeds, and then collide with a target or at opposite speeds collide with each other, and then are directed to a bubble chamber, where their tracks are photographed. Particle accelerators are mainly used to accelerate electrons, protons, and heavy ions. A distinction is made between resonant and nonresonant particle accelerators. The most promising are resonant particle accelerators (or resonators), in which particles are accelerated by a high-frequency electromagnetic field, and the accelerated particles move in resonance with changes in the field. There is also a distinction between linear particle accelerators, in which the accelerated particles move in a straight line, and cyclic particle accelerators (cyclotrons), in which they move in a circle. In modern cyclic particle accelerators, particles are accelerated to high speeds by repeatedly passing through an accelerating field. Synchronization of particle motion with a change in the field is performed by the autophasing method (synchrophasotrons, etc.). In the latest generations of particle accelerators, particles are collided in counter beams. Such particle accelerators are called colliders. As a result, it is possible to accelerate the particles to near-light velocity.

In 2012, the Large Hadron Collider at CERN discovered the new so-called **Higgs** particle, which was predicted by Peter Higgs in the framework of standard quantum theory in the middle of the last century. This is considered great luck, because thanks to this discovery quantum theory received another very important experimental confirmation. With the help of the Higgs particle it was possible, in particular, to explain the nature and magnitude of the masses of all elementary particles.

7.5. Quantum statistics

It has been repeatedly emphasized that the properties of macrobodies are determined by the properties of their microstructure. The

branch of physics that studies the properties of macrosystems, i.e. systems consisting of a huge number of identical particles (microstructures) and having therefore an equally large number of degrees of freedom, through the properties of these particles and their interactions, is called **statistical physics**. In section 1.3 we reviewed the basics of statistical physics from a classical point of view. Here we will also consider quantum statistics.

Usually, for simplicity, statistical physics studies isolated macrosystems. Let us recall, however, that strictly isolated systems do not exist in reality, since it is impossible to completely eliminate the interaction of the system with the external environment. Therefore, an isolated system is an idealization of reality. Realistically, we can speak only about a quasi-isolated system whose energy is continuously changing relative to some average value $\mathcal{E}$ with a variation $\Delta\mathcal{E} << \mathcal{E}$. From the quantum mechanical point of view, the system within the variation $\Delta\mathcal{E}$ passes from one quantum state, which is characterized by a discrete energy level $\mathcal{E}_i$, to another state with its energy level. These quantum states are called **possible microstates**. Each given macrostate corresponds to a huge number of microstates (energy levels) with a large multiple of their degeneracy. Since, from the point of view of the macrostate, the value of $\Delta\mathcal{E}$, within which all microstates are fit, is extremely small, such transitions remain virtually invisible. The set of all possible microstates for each given macrostate is called its **static ensemble**. Statistical descriptions of isolated systems are called **ergodic descriptions**.

As an example, let us consider a macrosystem in the form of some gas with a volume of 10^{-6} m^3 under normal conditions. Transitions from one microstate to another arise as a result of mutual collisions of gas molecules. Let n be the concentration of molecules in the gas, and let v be the average collision frequency, then the number of collisions per unit volume and time is

$$N_S = v \cdot n.$$

Since one mole of gas contains $N_A = 6{,}023 \cdot 10^{23}$ molecules, and the volume V_M of one mole is found from the ratio

$$V_M = \frac{\mu}{\rho}$$

where μ is the mass of one mole of gas;

ρ is the density of the gas,

then

$$n = \frac{N}{V} = \frac{\rho N}{\mu}$$

If we take hydrogen as a gas, then

$$\rho = 9 \cdot 10^{-2}\ \text{kg/m}^3\ ,\ \mu = 2\ \text{kg/kg-mole},\ \text{n} \approx 10^{25}\ \text{1/m}^{-3}$$

Substituting the values n and v into the formula for the number of collisions N_s, taking into account that under normal conditions the frequency of mutual collisions of gas molecules

$$v = 10^8\ \text{s}^{-1},$$

gives

$$N_S = 10^{33}\,\frac{1}{m^{-3}s}.$$

The change of microstates in the unit volume occurs in time

$$\tau_S = \frac{1}{N_S} = 10^{-33}\ s.$$

It is in principle impossible to catch a change in the system in such a short time, even with instruments that can be given the highest imaginable accuracy and resolution. Static physics accepts a postulate

called the ergodic hypothesis, according to which all microstates of an isolated system corresponding to the same macrostate, or all members of an ergodic ensemble, are equally probable. This means that in a sufficiently long time interval the system passes through all microstates, and in any averaged time interval it is in each of them equally often. It follows that the probability of macrostates of an isolated system is proportional to the number of microstates through which these macrostates are realized. Mathematically it can be written as follows

$$p_j = CN_j, \tag{7.149}$$

where p_j is the probability of the j-th microstate;

N_j is the number of corresponding microstates.

The value dS, proportional to the relative probability of a microstate dp / p, the Austrian physicist Ludwig Boltzmann called the **entropy of the system**. As a result of integrating the value of

$$dS = k\frac{dp}{p}$$

we get that

$$S = k \ln p. \tag{7.150}$$

This expression is called the **Boltzmann entropy formula**.

The set of probabilities p_j is called a **statistical distribution**, and the equation (7.150) is called a microcanonical distribution. Since in each microstate the system is characterized by a well-defined value of the microcharacteristic q_i, the ensemble average value of this characteristic Q can be written through the statistical distribution as follows

$$Q = \sum_{j=1}^{N_j} p_j q_j$$

This raises the question of how many microstates N_j are realized by the macrostate Q.

To answer this question, we consider the following simple example. Let's number the 10 balls and consider the number of ways to stack them in forward and reverse order. Then we get the most ordered sequence of all possible sequences. It can obviously only be realized by the combination $C_{10}^{10} = 1$. Let's break the order so that one of the balls, for example the 10th, is stacked arbitrarily. This can be accomplished in 10 ways by the combination $C_{10}^{1} = 10$. A smaller order corresponds to more options for its possible implementation. Further violation of the order for two balls, for example the 10th and 9th, can be realized already 45 ways by combinations $C_{10}^{2} = 45$, etc.

The biggest disorder (chaos) will occur if you allow arbitrariness in the stacking of half of all balls. This can be realized in 252 ways by the combination $C_{10}^{5} = 252$.

In this case, it is impossible to specify the rules of stacking, because in any option of it the condition of arbitrary stacking of 5 balls is satisfied. On the other hand, an arbitrary arrangement of balls makes the system most homogeneous. In statistical physics, such a state is called **statistical equilibrium**. Hence it follows that an ergodic ensemble always tends to static equilibrium, which is characterized by complete homogeneity, internal equilibrium and balance (see also sections 1.3 and 1.3.1).

Consider an isolated system in the form of an ideal gas and a medium surrounding it. Since, by definition, there are no interactions between particles of an ideal gas, we can apply to it the hydrogen-like approximation of quantum mechanics that we discussed in Section 7.4.1. The distribution of particles into one-electron quantum states is called **orbitals**. Let's assume that the ideal gas we are considering is in thermal and diffusion contact with the medium.

The external medium in this case is called the thermostat, and the gas is the working body or just a body. We denote the total number of particles in the system by n_0 and its energy by $\mathcal{E}_0$. We denote the number of particles and energy of the body by n and $\mathcal{E}$ respectively. Let us suppose that $n << n_0$ and $\mathcal{E} << \mathcal{E}_0$. The energy levels of the body is denoted by $\mathcal{E}_i$, and the multiplicity of degeneracy levels is denoted by q_i. The value q_i is called the **statistical weight of the levels**. *

Let the body be at the level of $\mathcal{E}_i$, then for the thermostat following is ture

$$\left.\begin{aligned} n_T &= n_0 - n \\ \mathcal{E}_T &= \mathcal{E}_0 - \mathcal{E}_i \end{aligned}\right\} \tag{7.151}$$

According to (7.149) for the system as a whole

$$p_{j_0} = C_1 N_{j_0} \tag{7.152}$$

Let us further assume, as is common in statistical physics, that

$$N = \prod_i N_j \tag{7.153}$$

This equation is fulfilled with great accuracy in weak interactions. Since for a body the number of possible microstates N can be taken as the statistical weight of its levels, then for the system as a whole according to equation (7.149)

$$p_0 = C_{q_i} N_T \tag{7.154}$$

where N_T is the number of possible microstates of the thermostat.

From the Boltzmann entropy formula applied to the thermostat, it follows that

$$N_T = e^{\frac{S}{k}}. \tag{7.155}$$

Let us further decompose the entropy value of the thermostat S ($n - n_0$, $\mathcal{E}_0 - \mathcal{E}_i$) into a series according to the small values of n and $\mathcal{E}_i$. Limiting this expansion to the first-order terms, we obtain

$$S(n_0 - n, \mathcal{E}_0 - \mathcal{E}_i) \approx S(n_0, \mathcal{E}_0) - \left(\frac{\partial S}{\partial n}\right)_{\mathcal{E}_0} n - \left(\frac{\partial S}{\partial \mathcal{E}_0}\right)_{n_0} \mathcal{E}_i \quad (7.156)$$

Let's take the following notation

$$\left(\frac{\partial S}{\partial \mathcal{E}_0}\right)_{n_0} = \frac{1}{T}; \left(\frac{\partial S}{\partial n}\right)_{\mathcal{E}_0} = -\frac{\mu}{T} \quad (7.157)$$

The values T and μ are called the statistical temperature and the chemical potential, respectively. In our case they are the characteristics of the thermostat.

Substituting equation (7.157) into (7.156) and into (7.155) gives

$$p(n, \mathcal{E}_i) = C_1 q(n, \mathcal{E}_i) e^{\frac{\mu n - \mathcal{E}_i}{kT}} \quad (7.158)$$

The constant C1 is determined from the normalization condition, according to which

$$\sum_{n,j} p(n, \mathcal{E}_i) = 1 \quad (7.159)$$

The final result is

$$p(n, \mathcal{E}_i) = \frac{1}{Z} q(n, \mathcal{E}_i) e^{\frac{\mu n - \mathcal{E}_i}{kT}}. \quad (7.160)$$

The value Z is obtained from (7.160) and is called the large statistical sum or the Gibbs sum

$$Z = \sum_n \sum_j q(n, \mathcal{E}_i) e^{\frac{\mu n - \mathcal{E}_i}{kT}}. \quad (7.161)$$

Equation (7.161) expresses the probability that a body in thermodiffusive contact with the medium (thermostat) contains n particles and has energy $\mathcal{E}_i$.

In the case when diffusion of the body can be neglected (there is no exchange of particles between the body and the medium), expressions (7.160) and (7.161) take a simpler form

$$p(\mathcal{E}_i) = \frac{1}{Z} q(\mathcal{E}_i) e^{\frac{\mathcal{E}_i}{kT}}. \quad (7.162)$$

$$Z = \sum_i q(\mathcal{E}_i) e^{\frac{\mathcal{E}_i}{kT}}. \quad (7.163)$$

*

The relation (7.162) is called the **canonical Gibbs distribution**, and (7.163) is called the **statistical sum**. Fig. 7.29 shows the dependence $p = f(\mathcal{E}_i)$. Here $\bar{\mathcal{E}}$ is the average energy of the system.

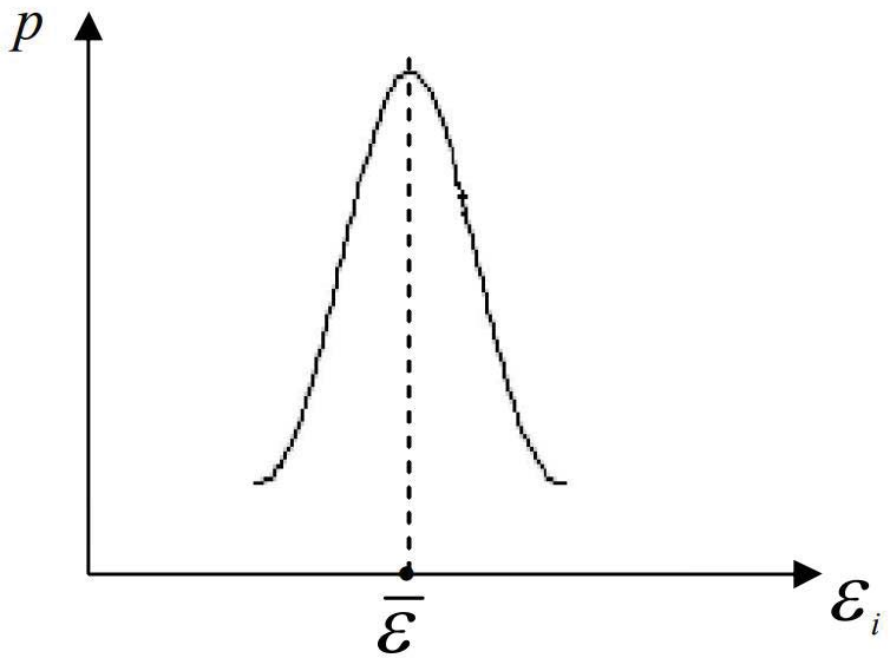


Figure 7.29.

As can be seen from the figure, the maximum likelihood of the canonical distribution corresponds to the average energy of the body, which is determined from the ratio

$$\bar{\mathcal{E}} = \sum_i \mathcal{E}_i p(\mathcal{E}_i) \quad (7.164)$$

The canonical Gibbs distribution can be applied to any orbital according to its definition, since the particles in this orbital are quasi-independent on the population of particles in all other orbitals acting as a thermostat.

Let us first consider particles whose wave function belongs to the class with odd symmetry, to which, therefore, the Pauli exclusion principle applies. There are two possible distributions for such particles:

- with an empty orbital, when the fill number and energy are zero;

- with an occupied orbital when the fill number is 1 and energy $\mathcal{E}_1 = \mathcal{E}$.

The average number of particles in a quantum state with energy $\mathcal{E}$ is called **orbital occupancy**. Extending relations (7.160) and (7.161) to the particles in question and taking into account that $n = 1$ and $q\,(1, \mathcal{E}) = 1$, we obtain

$$\left.\begin{aligned} p(1,\mathcal{E}) &= \frac{1}{Z} e^{\frac{\mu-\mathcal{E}}{kT}} \\ Z &= 1 + e^{\frac{\mu-\mathcal{E}}{kT}} \end{aligned}\right\} \tag{7.165}$$

For the orbital occupation we have, respectively

$$\bar{n} = \sum_k p(n_k) n_k \tag{7.166}$$

In our case k takes values 0 and 1, therefore

$$\bar{n}(\mathcal{E}) = p(0) \cdot 0 + p(1) \cdot 1 = p(1,\mathcal{E}) \tag{7.167}$$

Denoting, as is customary in quantum mechanics,

$$\bar{n}(\mathcal{E}) = f(\mathcal{E})$$

and substituting the equation (7.165) into (7.166), we obtain

$$f(\mathcal{E}) = \frac{e^{\frac{\mu-\mathcal{E}}{kT}}}{1 + e^{\frac{\mu-\mathcal{E}}{kT}}}$$

or

$$f(\mathcal{E}) = \frac{1}{e^{\frac{\mathcal{E}-\mu}{kT}} + 1} \tag{7.168}$$

The distribution (7.168) was established independently by Enrico Fermi and Paul Dirac and is therefore called the Fermi-Dirac statistic. Particles obeying the Fermi-Dirac statistics are called **fermions**, respectively. It has now been established that fermions include all matter particles that are part of atoms and atomic nuclei, as well as their antiparticles (electrons, positrons, nucleons and antinucleons, quarks and antiquarks), as well as neutrinos and antineutrinos that are not part of the atomic matter. Fermions also include all baryons and composite particles with half-integer spin.

Let us consider now a group of particles whose wave function belongs to the class with even symmetry. Let us apply the canonical distribution to an orbital containing n particles with energy $\mathcal{E}$. Since each orbital is characterized by one single state *j*, following is true for it

$$\left.\begin{aligned} q(n, \mathcal{E}_j) &= 1 \\ \mathcal{E}_j &= n\mathcal{E} \end{aligned}\right\} \tag{7.169}$$

In this case the summation over j is eliminated. According to (7.161)

$$Z = \sum_n e^{\frac{\mu n - \mathcal{E}}{kT}}. \tag{7.170}$$

Since the number of particles in the macroscopic volume of gas is enormous, we will consider the sum (7.170) as infinite. Then

$$Z = \sum_{n=0}^{\infty} \left[e^{\frac{\mu n - \mathcal{E}}{kT}}\right]^n. \tag{7.171}$$

Based on physical considerations, the series (7.171) should be considered convergent. Indeed, if we assume that $Z \to \infty$, then according

to equation (7.171) $p \to 0$, which does not make sense because the probability of all states cannot be zero. The convergence of the series is ensured unambiguously in all cases. If $\mu < 0$, then (7.171) is an infinitely decreasing geometric progression for which

$$Z = \frac{1}{1 - e^{\frac{\mu - \mathcal{E}}{kT}}} \tag{7.172}$$

Substituting (7.172) into (7.161) and then into the distribution function $f\ (\mathcal{E}) = \bar{n}$ gives

$$f(\mathcal{E}) = \frac{1}{e^{\frac{\mathcal{E}-\mu}{kT}} - 1} \tag{7.173}$$

*

The relation (7.173) is a symmetric distribution function for particles that are not subject to the Pauli exclusion principle. This distribution was first established by Bose and Einstein and is called the Bose-Einstein statistic. Particles to which this statistic applies are called **bosons**. It is now established that bosons include photons, gluons, gravitons, vector bosons, carriers of weak interactions, and all types of mesons. In addition, bosons include composite particles consisting of an even number of fermions.

The chemical potential of the gas μ is in turn a function of temperature. The level of the gas whose energy at T ≠ 0 is equal to the chemical potential is called the **Fermi level**. $\mathcal{E}_F = \mu\ (T\ )$ at temperature T ≠ 0 . The peculiarity of the Fermi level is that its occupancy in fermions, as follows from (7.167), is

$$f(\mathcal{E}_F) = \frac{1}{2} \tag{7.174}$$

It also follows from (7.167) that at T = 0 all orbitals for which $\mathcal{E} > \mathcal{E}_F$ are unfilled ($f = 0$), and for the orbitals for which $\mathcal{E} < \mathcal{E}_F$, $f = 1$ (see Figure 7.30).

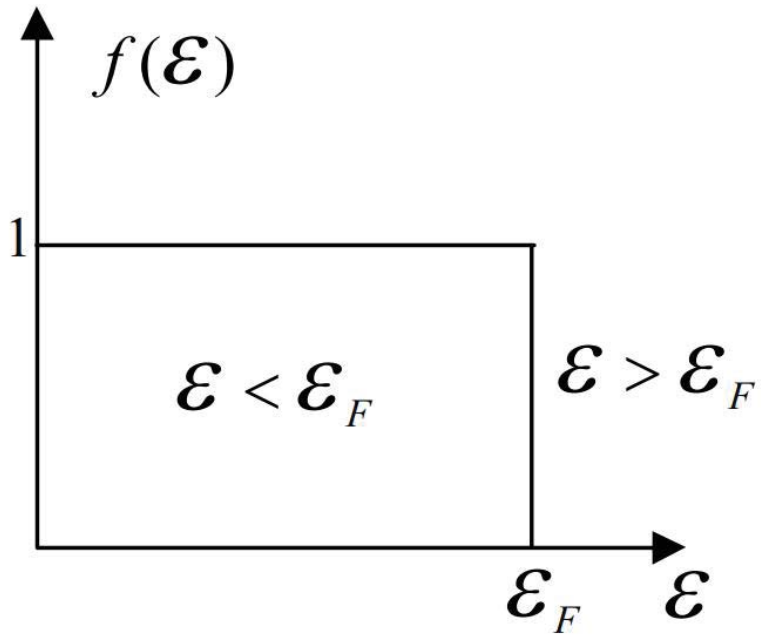


Figure 7.30

As an example of a fermionic gas, let us consider the electron gas in metals of the so-called ionic type. Inside the crystal lattice, the electromagnetic field of positive ions is compensated by the field of electrons so that the electrons are quasi-free, and the electron gas is close in its properties to an ideal gas, which obeys the Fermi-Dirac statistics.

This gas has a high energy and Fermi temperature. For silver (Ag), for example, the Fermi temperature $T_F = 64000$ K , for copper (Cu) $T_F =$ 82000 K, and so on. The electron gas state in crystals at absolute zero temperature is the basic quantum state with the lowest possible energy. However, the average energy of this state, while minimal, remains quite large and appears in the form of the kinetic energy of motion of electrons, whose speed is of the order of 10^6 m/s.

When a metal is heated, the energy of some small fraction of the electrons approaches the Fermi energy. However, on average, the energy of the electrons changes little with changes in temperature.

An interesting property of Bose gas is the Bose condensation. It consists in the fact that, at a certain temperature, gas particles move en

masse to the basic orbital. Some, a smaller part of them, remain in their orbits. It is as if the gas splits into two components with different properties. Helium, for example, at T = 4.2 K enters the liquid phase, although even after that the property of liquid helium is still described by Bose-Einstein statistics. Such fluids are called quantum fluids.

At T = 2.17 K, helium $^4_2$He undergoes Bose condensation and becomes two-component. The component whose particles are in the main orbital is practically devoid of viscosity. It is, as already pointed out, called superfluid. The properties of the second component are practically unchanged. Liquid helium at $T > 2.17$ K is called helium I, and at $T < 2.17$ K it is called helium II.

7.6. Quantum field theory

The fundamentals of field theory, including its classical and quantum aspects, are discussed in general terms in Sections 1.2.2 and 4. As follows from these sections, the **classical field theory** proceeds from the fact that space-time is formed by a set of constantly interacting discrete matter and continuously distributed physical fields in it, which connect all physical bodies through fundamental interactions. From the classical point of view, the discrete nature of matter manifests itself in its corpuscular structure, represented by a set of elementary particles.

All particles of matter, in turn, have some charge, the nature of which is determined by the type of interactions (gravitational, electromagnetic, etc.). Charges create physical fields continuously distributed in space-time around the particles, which have the property that at each point they act with some force on the charged particles of the same nature placed in it.

Classical physics, however, proved unable to answer questions related to the nature of physical fields and their sources, charges. It also

failed to explain why charges create a field around themselves and how physical fields are involved in the interaction between particles of matter.

7.6.1. Interaction Theory. Virtual particles and the physical vacuum

The difficulties encountered have found their solution in modern **quantum field theory**. From the point of view of this theory, it can be considered proven that the division of matter into matter and field has a **conventionally-classical** character and is justified only when considering **macro-processes**. First of all, it follows from the fact that in accordance with wave-particle dualism and consequent quantum-mechanical nature of micro-objects not only substance, but also field, has a clearly expressed discrete structure and is represented by a set of elementary particles. On the other hand, matter also does not have a purely discrete structure, and its particles are also characterized by wave properties. In other words, quantum mechanics has brought particles and fields much closer together. From its point of view, there is no fundamental difference between field particles, its quanta, bosons, and matter particles, fermions. Both have equally corpuscular and wave properties and differ from each other only by quantum statistics and values of some quantum numbers, including spin, parity, etc. So, for example, quanta of all fundamental fields are born and absorbed by discrete portions, particles which, like fermions, have definite momentum and energy. They differ from matter particles in that the wave function describing them is even, and the distribution function obeys not the Fermi-Dirac statistics, but the Bose-Einstein statistics (see Section 7.5).

It follows from the wave nature of all elementary particles of matter and field that they can be born and absorbed, as well as mutually transform into each other. This property is confirmed by many facts. Thus, an electron and a positron in the process of annihilation give rise to a pair of photons, and, conversely, photons can give rise to electrons and

positrons. Proton collisions produce pions, which in turn decay into muons and neutrinos. Protons, by emitting electrons, turn into neutrons, and neutrons, by absorbing positrons, turn into protons.

To describe wave processes, on the one hand, it is well known that continuous media (systems with an infinite number of degrees of freedom, see section 6.1) must be considered. On the other hand, the wave processes under consideration, in accordance with their corpuscular nature, must be subjected to **subsequent quantization**. The resulting field is reduced to a set of quantized excitation waves, which are also called **field quanta**. The theory of such systems is called **quantum field theory**. The mechanical analogue of quantized systems can be a set of interconnected oscillators that fill the entire space. In such a system, as is known (see section 6.1), collective oscillations can occur, each of which is characterized by its own natural frequencies. The connections between the oscillators contribute to the fact that the corresponding wave processes propagate through the system. In a quantum mechanical process, collective vibrations are quantized, and the resulting quanta are treated as particles with energy and momentum.

Quantum field theory is **relativistic**. It follows from the fact that due to mutual transformability of elementary particles in any system along with slow (nonrelativistic) particles of matter there are always fast (relativistic) massless particles, such as photons (massless are called particles with rest mass equal to zero).

In **quantum electrodynamics**, in particular, it is shown that the interaction between elementary electric particles is the result of their **exchange of photons**, which are emitted by one of the interacting particles and absorbed by the other. Thus electric charge in quantum electrodynamics is no longer a thing in itself, but is considered as a property of elementary particle related to internal, so-called **gauge symmetry** of field equations, discussed below. In addition, it is now shown in quantum field theory that the mechanism of all fundamental

interactions is identical and therefore the conclusions obtained in quantum electrodynamics apply to them.

Meanwhile, a free material particle, according to the law of conservation of energy of classical physics, can neither emit nor absorb a particle, a field quantum, in principle. In quantum theory, on the contrary, according to uncertainty relations, behavior of all real particles is **nondeterministic** and does not obey the laws of conservation of energy-momentum. In contrast to the classical principle of **causality**, to which all material objects must obey without exception, quantum objects, i.e. elementary particles, are characterized by a certain **freedom of choice** and are not subject to laws inherent to particles from their material nature.

Formally, this circumstance is unusual and causes some inconvenience. In this regard, physicists have agreed conditionally to assume that all elementary particles have a purely material nature, that is, they have assumed that their behavior is rigidly determined by objective physical laws, including the laws of conservation and causality.

In order to take into account the real properties of elementary particles, the theory introduced for each of them another particle that is subject only to the uncertainty relations, by means of which, therefore, exchange interactions take place. This particle was called a **virtual particle**. In fact, a virtual particle is not a real particle, but only a set of its possible unrealized states. According to the uncertainty relation a particle can take any state from this set. Its real state is only one of this set, which was realized randomly. It follows that the virtual particle became called a particle only by analogy with the material particle, although there is nothing in common between them. A virtual particle, by definition, should be considered non-material. The non-materiality of virtual particles is confirmed by the fact that their behavior, according to the uncertainty relations, is completely non-deterministic and causeless. For a very short time, they first appear randomly and then also disappear randomly and do not obey the so-called objective laws of Nature,

including the laws of conservation of energy and momentum. In this regard, they cannot in principle be **observed**, because they have time to disappear before any observer is able to detect them. In this sense, they are, by definition, **immaterial**. Thus, virtual particles in all their properties are **antipodes** of material particles and, accordingly, meet all the criteria of immateriality.

On the other hand, since material particles exist in space-time and virtual particles are immaterial, the latter can exist only **outside of space-time**. The above circumstance led to the need to introduce such an entity (conventionally, a medium), in which, by definition, there are **no material particles**, but only virtual particles that are continuously and randomly born and disappear. To ensure the law of conservation of charge, it was assumed that virtual particles arise and disappear in **pairs** as particles and antiparticles. This medium has been called the **physical vacuum**. Obviously, it is necessary to attribute immateriality (virtuality) to the physical vacuum as well.

In electrodynamics, the existence of the electromagnetic field (field strength is not equal to zero) is always accompanied by the birth of photons, but the absorption of photons, i.e. their disappearance, does not always lead to the disappearance of the electromagnetic field. This means that in the general case, that is, when the result obtained is extended to any fields, the field strength and the number of particles, its carriers, correspond to non-commutative operators and cannot be set with the same accuracy. Non-commutating operators are those whose result depends on the order in which they are applied. As an example of non-commutating operators in everyday life, the events that follow each other are entering and leaving a room. If we enter first and then leave the room, we end up outside the room. If you go out first and then enter the room, the result is just the opposite.

In quantum field theory it is proved that uncertainty relations apply only to pairs of physical quantities of non-commutative operators.

Since the number of material particles in the physical vacuum is always exactly zero, the field strength in it cannot be zero. The mentioned circumstance led physicists to the conclusion that it is necessary to consider the vacuum state not as a simple absence of a field, but as a **lower energy state** of quantized fields. This means that the physical vacuum represents one of the possible states of the field that can be detected experimentally. This approach, in turn, led to the consideration of the physical vacuum as a **special form of matter**. This was confirmed in practice when it was found that in the presence of physical vacuum the properties of material objects change. In particular, there is polarization of the physical vacuum, which manifests itself as screening by the vacuum of a part of the electric charge of a particle. In addition, there are so-called radiation corrections and level shifts in the form of superfine splitting of the radiation levels of the electron, etc. A clear contradiction has appeared. On the one hand, due to the absence of material particles in the physical vacuum by its definition, it cannot be material. On the other hand, precisely for the same reason, the physical vacuum should be considered a special form of matter.

This contradiction is considered within the framework of quantum field theory as some kind of a given. Meanwhile, it is easily eliminated if we consider that in the definition of the physical vacuum, as in the case of virtual particles, a certain arbitrariness is allowed. The point is that, if we follow exactly the definition of physical vacuum, it should be initially considered an **immaterial medium**, to which physical characteristics inherent to material entities, including field strength, energy, time, momentum, etc., are not applicable in principle. This means that the uncertainty relations are essentially inapplicable to it, although their formal use in describing the physical vacuum in order to better understand quantum processes is not excluded. The erroneous conclusion about the materiality of the physical vacuum arose precisely because the initial assumption is that the laws applying only to material objects, in this case the uncertainty relations, are applicable to it. In other words, they first

assumed that the physical vacuum is a material entity, and then discovered that it has, quite naturally, material properties.

From this we could conclude that the conclusion about the materiality of the physical vacuum is far-fetched. However, it is disconcerting that the material properties of the physical vacuum have indeed been discovered experimentally. This circumstance suggests that in the presence of matter and under certain conditions, the physical vacuum **materializes**. There is nothing unusual about this materialization, given that it is not at all limited to the transformation of intangible entities into material ones. Moreover, it is quite **natural** and applies to any non-material formations existing in the Universe, including virtual particles, monochromatic waves, ideas, categories of thinking activity, information, etc. Its essence lies in the fact that transforming factors, the existence of which is a consequence of the existence of corresponding non-material formations, are imposed on material objects. An example of such materialization is the phenomenon of **modulation**, which we discussed above and which is familiar to many in everyday life. It consists in the fact that on a material medium (e.g., light wave) is superimposed intangible information, such as information about the appearance of the object. This is done with a material signal created by the distribution of amplitudes and phases in the wave reflected from the object, the hologram. This hologram recorded, for example, on the retina corresponds exactly to the distribution on the surface of the body of amplitudes and phases in the wave reflected from the object, which corresponds to the individuality, in this case the visual image, of the appearance of the object in question.

7.6.2. Symmetry in quantum field theory

A special place in quantum field theory is occupied by **symmetry** of laws of nature, the notion of which emerged as early as in classical physics. Symmetry in Greek means **proportionality**, **equality**. A

distinction is made between geometrical symmetry (positions and shapes) and physical symmetry (physical laws and equations). Physical laws are called **symmetric** if they do not change under certain transformations to which the system can be subjected. In this case, the laws have symmetry or **invariance** with respect to the specified transformations. It follows from experience that physical laws are symmetrical with respect to:

- continuous global transformations of the **spatiotemporal group**;

- discrete **CPT transformations** of space-time;

- **permutation** of identical particles;

- local internal transformations that do not depend on spatio-temporal coordinates (**internal symmetry**).

The first group is a consequence of the homogeneity and isotropy of space-time, which lies outside the microcosm, as well as the equality of inertial systems (see Section 2.3). It is the result of continuous transformations of space-time by transferring, rotating the reference system in space as a whole, changing the origin of time and transition from one inertial system to another.

The second group of symmetries (CPT) includes discrete space-time transformations, including **spatial inversion** (P), which is reduced to changing the sign of spatial coordinates $\vec{r}$ to $(-\vec{r})$, **charge conjugation** (C), involving replacement of particles by antiparticles, and **time reversal** (T). CPT symmetry is a consequence of relativistic invariance (i.e. invariance with respect to Lorentz transformations) and locality of physical interactions. It defines the content of CPT theorem of quantum field theory, according to which equations of this theory are symmetric (invariant) with respect to CPT transformations carried out simultaneously.

The third group of symmetry means equality of elementary particles, which is a consequence of identity, i.e. indistinguishability of these particles.

The fourth group of symmetry means equality under certain conditions:

- individual groups of hadrons (isotopic or unitary symmetry in general);

- quarks of different colors and leptons (quark-lepton symmetry);

- fermions and bosons (supersymmetry);

- fundamental interactions whose source are conserved charges (gauge symmetry).

Isotopic invariance refers to particles participating in strong interactions, **hadrons** that form **isotopic multiplets**. An isotopic multiplet is a group of particles with different electric charge but approximately the same masses and exactly the same spin, isotopic spin, internal parity, baryonic charge, strangeness, charm and beauty (see Section 7.4.10).

The total number of values that the isotopic spin projection (7.140), electric charge (7.141), and hypercharge (7.142) can take is determined by the number of multiplet members.

For duplets, such as nucleons,

$$I = \frac{1}{2}; I_Z\left(+\frac{1}{2}; -\frac{1}{2}\right); Q(Q_p = +1, Q_n = 0)$$

$$Y = 1\ (B = 1,\ S = C = b = 0).$$

For triplets, e.g., π-mesons, the hypercharge and electric charges, respectively, are found from the relations

$$I = 1;\ I_z\ (+1;\ -1;\ 0);\ \mathrm{Q}\ (+1;\ -1;\ 0).$$

$$Y = 0\ (B = S = C = b).$$

Moving from one particle to another within a given multiplet does not, as we see it, change the value of its isotopic spin, but it changes its projection, which is mathematically equivalent to rotating the vector of isotopic spin in some isotopic space. This corresponds to symmetry group SU(2) if the number of varieties (flavors) of quarks included in these particles equals two of 3 flavors, to group SU(3) if the number of varieties (flavors) of quarks equals 3 of 3 flavors, to group SU(n) if the number of varieties (flavors) equals 3 of n. It follows that nucleons and pi-mesons composed of quarks of two flavors (u,d), as well as sigma-hyperons and K-mesons, are hadrons with SU(2) symmetry, while other baryons are hadrons with SU(3) symmetry or higher. Indeed, studies have shown that there are unions of hadrons with higher multipletting. This means that isotopic symmetry is part of a more general, **unitary**, symmetry of SU(n) groups, which unites hadrons, which are a combination of 3 of n varieties (flavors) of quarks, into a family of multiplets with a large number of members.

Both isotopic and unitary symmetries are imprecise. Weakly broken by electromagnetic interactions, the isotopic symmetry includes hadrons with slightly different masses. The unitary symmetry is broken much more strongly. This is due to the fact that the masses of s, c, and b quarks included in hadrons united by unitary symmetries differ significantly from the masses of u and d quarks forming isotopic symmetries.

Strong interactions are also symmetric with respect to the color transformations of quarks, which constitute the exact group of 3 colors - SU(3) - C.

In the high energy region, there is a similarity between the electrically weak interaction of quarks and leptons. This suggests that there is a deeper quark-gluon symmetry underlying the great unification.

Supersymmetry unites fields whose quanta are both bosons and fermions. In the real world, supersymmetry is broken. There are no bosons and fermions of equal or at least close mass and no fields whose quanta are fermions.

According to the **Noether's theorem**, each symmetry transformation corresponds to a value that is **conserved**. For example, the laws of conservation of momentum, angular momentum, and energy follow from the symmetry of the space-time group. Symmetry with respect to the spatial inversion transformation leads to the preservation of spatial parity. The charge conjugation preserves the charge parity in the processes of strong and electromagnetic interactions, when the truly neutral system remains unchanged when the signs of the charges change. From the symmetry with respect to the reversal of time (changing the sign of time) follows the preservation of the form of the equations of motion and the reversibility of elementary processes. From symmetry with respect to local gauge transformations follow the laws of conservation of charge, from isotopic invariance follow the conservation of isotopic spin in strong interaction processes, and so on.

The symmetry of laws with respect to the corresponding transformations and the consequent preservation of physical quantities characterizing the system reduces its orderliness, level of organization, and functionality. The prevalence of symmetry in Nature is a consequence of the **universal law of symmetry**, according to which all material aggregates tend to symmetry, that is, to disorder. It is necessary to distinguish global, averaged in time and space symmetry from local symmetry. For example, the local homogeneity and isotropy of space-time, would turn the Universe into a giant vacuum desert. Fortunately, as shown above, at one stage of the evolution of the Universe, its

structurization occurred, which broke its local symmetry while maintaining, on average, homogeneity and isotropy. The same can be said about the symmetry of the material totality. It follows from the law of symmetry that the whole totality must, tending to symmetry, level out. However, this does not happen, because Nature has taken care to ensure that simultaneously with the emergence of new orders, its defense mechanisms would normally emerge to ensure the preservation of orders and identities for a long time.

Gauge symmetry, as stated above, belongs to the class of internal symmetries of field equations. It characterizes **sublocal symmetries** related to the properties of elementary particles described by parameters that depend directly on the coordinates of space-time points. Calibration symmetry reflects the fact that there are some **conserved** physical quantities characterizing elementary particles and generalized as **charges**, which, while tending to conservation, are simultaneously **sources of corresponding fields**. Charge, therefore, is not a structural unit, but a physical characteristic of a particle that presumably determines its internal structural order. The gauge symmetry in this case corresponds to a conserved charge (conserved internal order). It can be seen in the sense that it resists any attempt to change the internal order by external factors and seeks to restore it.

From the mathematical point of view, the transformation associated with the violation of the internal order is understood as a change in the wave function of a particle by multiplying it by the phase multiplier containing the value of the charge z, $-e^{jbz}$, which does not change the absolute value of the wave function value. Here b is an arbitrary coefficient. This means rotating the wave function by some angle bz in the isotopic phase space. The angle of rotation in isotopic space is a function of coordinates in space-time. When rotated by the phase angle, the equations of motion describing the interactions of the particles **change**. To ensure that the equation of motion remains unchanged, terms expressing the corresponding **vector fields** should be

added to it to compensate for these changes. These fields are called **gauge fields**.

In quantum electrodynamics, for example, it is proved that the **electromagnetic field** is the gauge (compensating) field for transformation of the electron wave function. The role of isotopic space in this case is performed by the complex phase space of the wave function.

The so-called **Yang–Mills fields**, whose quanta are the gluons that determine the strong interactions between quarks, are also gauge fields.

Quanta of gauge fields are called **gauge particles**. Gluons, for example, are gauge particles of strong fields, and photons are gauge particles of electromagnetic fields. It is known that gluons and photons are characterized by a rest mass equal to zero and spin equal to one. It turned out that these two properties are inherent to any gauge particles. For example, in the case of weak interactions, the conserved interaction constant (weak charge) is the isotopic spin corresponding to the local group SU(2). From the point of view of gauge symmetry theory, compensation for isotopic spin changes requires the introduction of three gauge fields. The sources of these fields are the projections of weak isospin and hypercharge, and their gauge particles, according to this statement, are three massless bosons, positive, negative and neutral, with spin equal to 1.

The problematic nature of the weak interaction mechanism thus constructed is that this interaction is short-range and must be carried by massive quanta of fields in connection with it, which, however, contradicts the original assumptions of gauge theory.

This contradiction is resolved with the help of the hypothesis of the English physicist Peter Higgs about **spontaneous symmetry breaking**. Spontaneous symmetry breaking occurs in the general case

when the state of systems that are described by equations of motion having symmetry is energetically disadvantageous. This means that when the system gets into conditions in which a state with lower energy can be realized, it, under certain energy ratios, spontaneously deviates from symmetry and goes to this state, increasing its level of orderliness. This disturbance is accompanied by a change in the charge of the particles, as well as a change in their structural ordering and a corresponding increase in mass. In this case, in particular, the birth of massive particles from massless particles and the self-generation of matter is possible.

To clarify the situation, let us consider the mechanical analogue of spontaneous symmetry breaking.

For example, let a ball fall along the axis of a cylinder with a spherically convex bottom. As a result of symmetric motion and subsequent falling to the bottom of the cylinder the ball will roll to its wall, i.e., it will spontaneously move to one of many possible (virtual), more ordered asymmetric states, but with less energy.

The level of symmetry is determined by the ambient conditions - temperature, pressure, concentration or generalized - energy of motion. In the initial Universe, according to Gamow's hot model, a huge energy was released as a result of the big bang, which was measured by the temperature at 10^{32} K (see section 1.6.2.2, and section 1.7, problem 1). Under these conditions, supersymmetry prevailed in the Universe, all particles did not yet possess wave-particle duality, had a point structure and behaved as perfect corpuscules. This means that they had integer spin equal to 1, rest mass equal to zero, moved at the speed of light, without interacting with each other.

According to the Higgs hypothesis, immediately after the big bang, an exotic field (called the Higgs field) arose, which had a braking effect on quarks and leptons equivalent to their deviation from supersymmetry. As a result, they structured themselves by reducing their

velocities, and gained charge and mass. This is how fermions were born in the Universe, separated from the remaining bosons, which did not interact with the Higgs field. At the same time, as a result of the interaction of the born fermions, the corresponding strong, weak, and electromagnetic gauge fields emerged. In addition, the birth of massive particles led to a slowing of the rate of time flow, intervals deviated from linearity, space-time acquired curvature, deviating from local symmetry, and under the action of the resulting gravity, structured.

According to another version, the spontaneous deviation of the initial Universe from supersymmetry arose under the influence of the materialized physical vacuum, which possesses, as mentioned above, a minimum of energy.

7.6.3. Fundamentals of Quantum Electrodynamics

Quantum electrodynamics is the most important part of quantum field theory. In particular, it considers the electromagnetic field as a gauge field with undisturbed symmetry and conserved charge. The gauge particle of this field is a photon, i.e. a massless particle with spin equal to one. Another peculiarity of quantum electrodynamics is that, as already mentioned, it interprets the interaction of electrons as the result of exchange between them by **virtual photons**. The intensity of electromagnetic interactions is characterized by a dimensionless bonding constant equal to

$$\alpha = \frac{e^2}{4\pi\varepsilon_0\hbar c} = \frac{1}{137}$$

Let us consider in more detail the action of the gauge transformation in quantum electrodynamics. Let in the general case be given two particles, which differ from each other only by their electric charge e and are described, respectively, by the wave functions

$\psi_p(\vec{r},t)$ and $\psi_n(\vec{r},t)$

The mentioned functions in this case can be considered as functions describing different charge states of the same particle. This particle can obviously be described by some complex function $\psi^*(\vec{r},t)$, which results from the superposition of the functions ψ_p and ψ_n. From the mathematical point of view, these functions can also be considered as components of some vector given in the space ψ_p and ψ_n, which in this case is called **isotopic**. This vector can be characterized not only by its components, but also by the rotation angle φ in isotopic space, which is determined by the magnitude of the charge. Since ψ_p and ψ_n are functions of coordinates and time $(\vec{r},t)$, then for the rotation angle $\varphi(\vec{r},t)$. A rotation by such an angle is called a **gauge transformation**. This transformation allows us to pass from one state of a particle to another, and the rotation of the vector in isotopic space, characterizing this transition, can be regarded as equivalent to some superposition of the functions ψ_p and ψ_n.

The laws describing the behavior of the particle in question are obviously independent of its state (in this case charge state). Therefore, it can be argued that these laws are not violated by the gauge transformation. This means that they are **invariant** with respect to the angle of rotation of the vector. This invariance is relativistic, because the laws of nature are invariant in this case with respect to the Lorentz transformations. In other words, these laws do not depend on the local choice of one or another superposition of the functions ψ_p and ψ_n.

It follows from the relativistic invariance of the laws of motion that the wave function $\psi^*(\vec{r},t)$ describing it changes both with time and with changes in coordinates.

This means that when the complex vector of the wave function in complex phase space is rotated by an angle $\varphi(\vec{r},t)$, the equations describing the behavior of the particle will acquire an additive

proportional to the change in angle φ as a function of $\vec{r}$ u t. This additive in Lorentz transformations changes as a four-vector.

Since, however, the equations of motion are invariant with respect to the Lorentz transformations, in order to compensate for this addition, we should introduce functions describing vector gauge fields that would acquire the same addition, but with opposite sign, when rotated. These fields prevent changes in the charge and leave it constant. In other words, gauge fields give rise to interactions of particles, due to which gauge symmetry is ensured.

In quantum electrodynamics, the real and imaginary parts of the **electron** wave function play the role of components of the complex vector

$$\psi_e^*(\vec{r},t) = \psi_e(\vec{r},t)e^{jeφ(r,t)} \qquad (7.175)$$

where ψ_e* is the imaginary part of the electron wave function;

e is the charge of the electron.

The isotopic space is the plane of the complex variable, where the real and imaginary parts of the function ψ_e* are plotted along the axes.

The relativistic Schrödinger differential equation for the wave function of the free electron, taking into account the requirement of relativistic invariance, was obtained by Dirac in 1928. Analysis of this equation showed that the gauge field of the electron wave function in this case is described by Maxwell's electromagnetic field equations. The calibration particle of this field is a massless particle with spin 1, equivalent to a photon.

Quantum electrodynamics has received a large number of experimental confirmations within a wide range of spatiotemporal intervals, from cosmic at 10^{18} to 10^{20} m, to microscopic intranuclear at 10^{-15} m.

Quantum electrodynamics, in particular, describes the processes considered earlier with high accuracy almost completely coinciding with the experiment. These can include thermal radiation processes (see item 7.2.1.2), Compton effect (see item 7.2.2), braking (X-ray) radiation (see item 7.4.4.3) and electromagnetic interaction processes (see items 6.5.7, 6.6.3) at any of the above levels. They also include such specific quantum processes as polarization of physical vacuum, scattering of light on light, birth of pairs of antiparticles in strong electromagnetic and gravitational fields and opposite processes of annihilation of particles, etc.

The specific processes of quantum electrodynamics are outlined in the following sections of the chapter.

7.6.4. Fundamentals of Quantum Chromodynamics

Quantum chromodynamics is a quantum theory of strong interaction of quarks and gluons, modeled on quantum electrodynamics based on gauge symmetry theory. In quantum chromodynamics, in contrast to quantum electrodynamics, instead of a single charge, which is the source of the gauge electromagnetic field, there are three color charges belonging to three quarks of the corresponding colors. In connection with this, in the color calibration theory, during the transition from one state to another, the gauge transformations change not only the phase, but also the color, i.e., the complex wave function is given by three matrix equations

$$\psi^{\alpha}(x) \to \sum_{\beta} e^{j\varphi_{\beta}^{\alpha}\psi^{\beta}(x)},$$

where the indices α and β correspond to the three possible color values of quarks.

This means that instead of a single rotation angle φ (x), all color states are enumerated by superposition of rotations by nine possible

angles $\varphi^{\alpha}{}_{\beta}(x)$, where $\alpha, \beta = 1, 2, 3$. One of these angles corresponding to the matrix sum

$$\sum_{\alpha} \varphi_{\alpha}^{\alpha}$$

is independent of time and coordinates. Therefore, in order to compensate for changes in the equations of motion, it is necessary to introduce eight gauge (color) fields instead of nine. These fields are usually called gluon or **Yang-Mills fields**. Quanta of these fields are color, i.e. charged massless gauge particles, gluons, with spin equal to 1. The exchange of virtual gluons leads to a strong interaction. Since gluons, like quarks, are charged, they themselves generate gluon fields. Emitting and absorbing virtual gluons, they interact both with each other and with quarks. This means that the gauge field equations for gluons are nonlinear. Fields of the electromagnetic type, which are described by linear equations, as well as their gauge theories, are called **Abelian fields**. Gauge fields that are described by correspondingly nonlinear equations and their theories are called **non-Abelian fields**.

The bonding constant of strong interactions is also determined, as in the case of the electromagnetic field, by the formula

$$K_s = \frac{g^2}{4\pi\varepsilon_0\hbar c} = 14,$$

where g is the value of the effective charge of the strong field.

This means that the intensity of strong interactions is more than 100 times greater than the intensity of electromagnetic interactions.

In ordinary matter under normal temperature conditions, strong interactions do not cause any processes, and their role is reduced only to the creation of strong bonds between quarks within nucleons and between nucleons in nuclei. This is because strong interactions are short-lived.

Their radius of action is only 10^{-15} m and does not go beyond the atomic nucleus. This essentially distinguishes them from electromagnetic interactions, which are responsible for the absolute majority of processes in matter. Indeed, the radius of action of electromagnetic interactions is theoretically infinite. Practically, the bipolarity of electric charges, the number of which is the same in atoms, neutralizes the effect of the electromagnetic field already at the molecular level, i.e. at distances of the order of 10^{-9} m. Nevertheless, these are the ones that determine all atomic and molecular processes that play a leading role in Nature.

Under experimental conditions in which nuclei or nucleons with sufficiently high energy collide, nuclear reactions occur, resulting in radioactivity and fusion processes, which are accompanied by the release of high energy and create all energy processes in the universe. Starting at energies of a few hundred megaelectronvolts, strong interactions give birth to pions, which ensure the bonding of nucleons within the atomic nucleus, and at even higher energies, to the birth of strange, charmed, and beautiful particles.

The small distances at which strong interactions appear follow from the logic of chromodynamics. Suppose, for example, that when virtual gluons are exchanged, one of the nucleon quarks gets enough speed to leave the nucleus. As a result, as indicated, a quark-antiquark pair is born. The antiquark of this pair combines with one of the nucleon quarks, discoloring it and turning it into the virtual qq* meson responsible for the intranuclear forces. The remaining quark from this pair joins the two quarks of the nucleon, discoloring them. As a result, the nucleon returns to its initial state. This prevents a quark from escaping from the nucleus and passing to a free state, as well as from violating the nucleon's stability. In other words, any attempt of a quark to escape from the nucleus brings the whole system back to a stable initial state and makes long-range interaction impossible.

7.6.5. Weak interactions

The name of this interaction follows from the fact that its intensity is much lower than the intensity of strong and electromagnetic interactions. In practice, the strength of interactions is estimated by the measurable rate of induced processes at the same energies. The energy at which velocities are estimated is usually taken to be 1 GeV, which is typical for elementary particle physics. At this energy, the minimum rate of processes such as radioactive decay under the influence of strong interactions is 10^{-23} s, rate of electromagnetic processes is 10^{-21} s, and rate of so-called weak processes such as β-decay is 10^{-10} s. The intensities of these interactions also correlate with each other in the same proportion. The binding constant for weak interactions is respectively 10^{-15}

$$K_w = \frac{q_w^2}{4\pi\varepsilon_0\hbar c} \approx 10^{-15},$$

where q_w is the effective charge of the weak field.

The intensity of weak interactions increases rapidly with increasing energy. For example, the low-energy process, the β-decay of the neutron, which occurs with an energy release of only about 1 MeV, lasts almost 10^3 s. By comparison, the β-decay of a Λ-hyperon lasts for 10^{-10} s, that is, it occurs with an intensity 10^{13} times greater, while the energy release is increased by a factor of only 100. At a significant energy increase of 10^5 times (up to 100 GeV), which corresponds to distances of colliding particles much smaller than the so-called Compton wavelength, equal to $2 \cdot 10^{-18}$ m, the intensity of weak interactions increases dramatically and approaches the intensity of electromagnetic interactions.

The radius of action of weak interactions is very small and is on the order of $2 \cdot 10^{-18}$ m. This means that weak interactions are concentrated deep inside the atomic nucleus and affect intranuclear processes. Despite their extremely small magnitude and short duration,

weak interactions play an enormous role in internuclear processes, as well as in the processes associated with the functioning and evolution of stars and, consequently, the Universe as a whole.

The processes of thermo-radioactive fusion and β-decay of radioactive atomic nuclei, common in Nature, are associated with weak interactions. The former are accompanied by the emission of neutrinos (antineutrinos) and serve as the main source of energy in the Universe. The second play an important role in the mutual transformations of protons and neutrons within nuclei, which determine the natural radioactivity of the latter, the dynamics of their functioning, and the very possibility of their existence. In addition, they ensure the emergence in Nature of a certain excess of electrons (positrons) and neutrons, as well as the circulation of neutrinos, without which the Universe would be doomed to death. In space, weak interactions do not only lead to fusion. On the one hand, they cause the explosion of so-called supernovae and their evolution up to black holes. On the other hand, these explosions result in the formation of interstellar matter that serves as raw material for second-, third-, and so on-generation stars, to which our Sun belongs.

The theory of β-decay was constructed by Fermi in 1934 by analogy with the theory of electromagnetic processes. Fermi based his theory on the contact interaction of two so-called weak currents, which are formed by four fermions, namely, the proton (p), neutron (n), electron (e), and neutrino (ν). These currents, according to Fermi, are the result of birth and annihilation of these particles in the process of their mutual transformation. The first current, called the nucleon current, converts the neutron to a proton. It is denoted as $p\bar{n}$ through the proton birth p and neutron annihilation $\bar{n}$ operators. The second current, called the lepton current, translates the neutrino into an electron and is denoted by $e\bar{\nu}$ through the electron creation e and neutrino annihilation operators, respectively.

These operators are noncommutative and form non-Abelian groups. By analogy with the isotopic symmetry of strong interactions, they lie at the base of the symmetry of the SU(2) group. It also follows that the electron and antineutrinos escaping from the radioactive nucleus during β-decays are not contained in it, but arise due to weak interactions during intranuclear transitions of nucleons. This process is analogous to the emission of photons by excited atoms.

Weak currents, unlike electromagnetic current, change the charge of particles and are therefore called charged (electromagnetic currents, as well as any other currents that do not change the charge of particles, for the same reason are called neutral). Subsequently, the theory of weak interactions was improved by the American physicists Murray Gell-Mann, Robert Marshak, Richard Feynman, and E. C. George Sudarshan, who in 1957 proposed a universal, so-called V-A theory. According to this theory, the total weak charged current is the sum of lepton and quark currents. Subsequently, it turned out that the lepton current consists of three components, which are created by the conversion of three varieties of neutrinos (electron, muon, and tau lepton) into, respectively, electron, muon, and tau lepton. In turn, the quark currents form, in total, a charged hadronic current consisting of nine quark summands. Finally, Murray Gell-Mann, his American colleagues Sheldon Glashow and Steven Weinberg, as well as Pakistani physicist Abdus Salam, within the framework of quantum field theory and gauge symmetry theory, suggested that the weak interaction is not contact. It, like other fundamental interactions, is carried by bosons, particles of gauge fields. In addition, the weak interaction involves not only components of weak charged currents, but also components of neutral currents, caused both by mutual transformations of each of the neutrino varieties and quark varieties. Interactions of charged currents, according to the above hypothesis, are carried by charged bosons (denoted by W^+ and W^-, respectively), and interactions of neutral currents are carried by neutral boson Z^0 and photon.

The triplet of bosons called intermediate, W^+, W^-, and Z^0, by analogy with strong interactions, is characterized by a conserved isotopic spin equal to 1, as shown above, with components +1, –1, and 0 and, respectively, the symmetry group SU(2). By the same analogy, we can take the isotopic spin of the photon forming the singlet to be 0, and its components to be +1 and –1. Accordingly, the symmetry to which the electromagnetic field obeys is characterized by a single charge and is described by the group U(1).

7.6.6. Fundamentals of quantum unified field theory

From the time of the ancient thinkers to the present, scientists have sought to reduce the World to a single beginning. It is this intention that has been the source of the materialistic concept of a unipolar world that has existed for about three millennia. Even Democritus laid a single brick in the foundation of the Universe in the form of the smallest particle of matter, which he called the atom. The atomistic model was subsequently brilliantly confirmed by molecular-kinetic theory and experiment. The fabulous successes of classical mechanics have created the impression that a huge variety of natural phenomena, including quite complex thermal and electromagnetic processes, are reduced to a relatively simple mechanical motion. This raised hopes that unified physics would appear in the very near future. However, the further development of science has not lived up to these hopes. It turned out that the processes taking place in the microcosm have very little in common with mechanical motion, which in some cases loses any physical meaning here at all. The past century was also marked by the rapid growth of discoveries of a variety of new elementary particles that make up the atom and the atomic nucleus. It seemed that the dream of a unified physics had finally collapsed. However, by the middle of the 20th century, due to the successes of quantum field theory, the situation changed dramatically again. First of all, it was shown that all the various interactions that exist in Nature can be reduced to just **four fundamental** ones. Work on

classification of elementary particles, started in this connection, revealed deep similarity between them and allowed to suggest that this similarity was based on single structure of matter, the same single brick of the Universe, which was predicted by Democritus 2500 years ago. The subsequent quark theory and theories of isotopic, unitary, and gauge symmetry, as well as the supersymmetry theory, allowed reducing more than a few hundred elementary particles to several varieties of leptons and a small number of multiplet hadrons with similar or close properties. Recently, the theory of combining electromagnetic and weak interactions has been created, and at present the theory of the great unification of weak, strong and electromagnetic interactions is being created, in which it seems to be possible to include gravitation as well.

Let us consider in more detail the theory of unification of fundamental interactions. It is now established that the division of interactions into 4 types, differing in their intensity, radius of action, concentration at different structural levels and specificity of carriers, is justified only at not too high energies acting at the upper levels. It turned out that when penetrating deep into the nucleus of the atom and the corresponding increase in energy, the interaction constants are compared, and the differences between the subnuclear particles and types of interactions are smoothed out. This led to the assumption of a single physical nature of all kinds of fields and elementary particles. At the beginning of the last century, the hypothesis of wave-particle duality revealed a profound similarity between particles of matter and fields. Then it turned out that the huge variety of elementary particles is conditional and that in fact they are all combined into three classes of common physical nature, namely, hadrons, composed of quarks, leptons and field quanta, bosons. On closer examination, it became clear that in the field of ultrahigh energies there is a further leveling of elementary particles and their interactions. This allows us to conclude that one of the most important attributive properties of the material world is associated with its unstoppable desire to preserve symmetry, to destroy structural

ordering and individuality, the transition of matter and field in the depths of atomic nuclei to a vacuum-like state. The distinction observed in practice between field and matter, on the one hand, and the elementary particles representing them, on the other hand, as well as the differences between types of fundamental interactions, is the consequence of relatively stable gauge fields operating at the upper levels.

The common nature of the interactions manifests itself only at very high energies, higher than 10^3 GeV. At these energies, the difference between the weak and electromagnetic interactions disappears and a single electrically weak field emerges. At even higher energies, on the order of 10^{16} GeV, the constants of the electrically weak and strong interactions are compared and mutual transformations of quarks and leptons become theoretically possible. Finally, supersymmetry, which should take place at even higher energies, on the order of 10^{19} GeV, unites into single groups fermions and bosons, including bosons with spins 3/2 and 2, which are attributed to quantum gravitational field particles, hypothetical gravitinos, and gravitons respectively.

As already mentioned, the standard theory of electrically weak interaction was created in the 1960s by American physicists Sheldon Glashow and Steven Weinberg, as well as Pakistani physicist Abdus Salam. According to this theory, the weak interaction, like other fundamental interactions, occurs in an exchange way. It is transported by virtual heavy intermediate vector bosons W^+, W^-, and Z^0 and a massless photon A with spins equal to 1, which are quanta of vector gauge fields of weak interaction and gauge field of electromagnetic interaction. The sources of these fields are projections of weak conserved isotopic spin. In this case, the field of photon A and the fields of vector bosons B^0, W^0 and Z, are related by the following linear relations

$$A = B^0 \cos\beta + W^0 \sin\beta \qquad (7.176)$$

$$Z = -B^0 \sin\beta + W^0 \cos\beta \qquad (7.177)$$

where β is determined from the equation

$$\tan\beta = \frac{g'}{g}, \qquad (7.178)$$

where g' is the weak conserved hypercharge constant;

g is the effective weak charge.

The relationship between the electric charge e and the weak charge g is given by the expression

$$e = g \sin\beta \qquad (7.179)$$

Occurrence of mass in intermediate bosons occurs with spontaneous breaking of gauge symmetry so that the rest masses of the W^+ and W^- bosons are 80 GeV/c^2, of the Z boson is 90 GeV/c^2, and the rest mass of the photon is zero, respectively. This theory of electrically weak interaction has found brilliant confirmation in practice. This happened after vector bosons were discovered at the end of the last century, whose characteristics fully matched those predicted. In addition, the connection between photons and vector bosons indicated in this theory was confirmed.

Following the creation of the theory of electrically weak interactions, it has also been observed that in the high energy range, gauge theories predict an increase in the electrically weak interaction constant, on the one hand, and a decrease in the strong interaction constant, on the other. Extrapolation of these predictions to the region of ultra-high energies, made it possible to assume that at the level of energies 10^{14} to 10^{16} GeV should have equality of all three constants. This made it possible to combine the electrically weak and strong interactions within the framework of the **Grand Unified Theory** (GUT). In this case photons, gluons, and intermediate vector bosons should form a multiplet of quanta of gauge fields of a single gauge quark-lepton symmetry. Moreover, since the specified energy is close to the value of the so-called Planck mass, equal to 10^{19} GeV/s^2, there is hope that within the framework of the quantum theory of gravitation it will be possible to

combine gravitational interaction with the Grand Unified Theory. Experimental verification of these assumptions is currently problematic, since it is not possible to obtain energies of the required magnitude in the foreseeable future. At the same time, according to Gamow's so-called hot Universe model, these energies took place in the initial Universe, at a point in time close to the Big Bang. This explains the processes that took place in the initial Universe. At the same time, combining quarks and leptons into one multiplet leads to quark-lepton transitions and proton decay. Calculations give as yet unconfirmed by experiment the proton lifetime of the order of 10^{30} years. In principle, this could explain the charge asymmetry observed in the Universe (the inequality of the number of particles and antiparticles) and the presence of neutrino rest mass. The theory also predicts the existence of magnetic monopole-type solutions, the existence of which also follows from the Dirac equation obtained back in 1927. The mass of predicted **magnetic monopoles** is so large that it is, unfortunately, not yet detectable by laboratory means.

7.6.7. Perturbation theory

The perturbation theory method used to calculate quantum field theory processes was proposed by the American physicist Richard Feynman. This method, also called the Feynman diagram method, consists in taking into account, step by step, an increasing number of acts of interaction of free particles, which are treated as point objects. Examples of Feynman diagrams are shown in Figures 7.31, 7.32, and 7.33.

The graphical symbol of particle propagation in Feynman diagrams is conventionally assumed to be some line, and real particles, fermions, are compared with a solid line and field quanta with a dashed line. The arrows on the lines indicate the direction of propagation of the particle. Antiparticles are matched with a solid line with an arrow showing the direction of propagation opposite to that of the

corresponding particle. Virtual particles are considered to be particles that are born and then absorbed at intermediate stages. In Feynman's diagram, it correspond to the lines connecting the meeting points, which are denoted by Arabic numerals. In the first, second, and so on approximations, one-time, two-time, and so on acts of interaction between different particles are taken into account. The number of acts of interaction is called the order of the diagram. The time axis is considered to be directed to the right and is not directly depicted in Feynman's diagrams. All diagrams are composed of the simplest elements, the vertex parts, consisting of three lines converging at the meeting points and representing either emission of a particle or its absorption, or birth by the field quantum of a particle-antiparticle pair, or, finally, annihilation of the pair. A second-order diagram describing the scattering of a photon on an electron is shown in Figure 7.31.

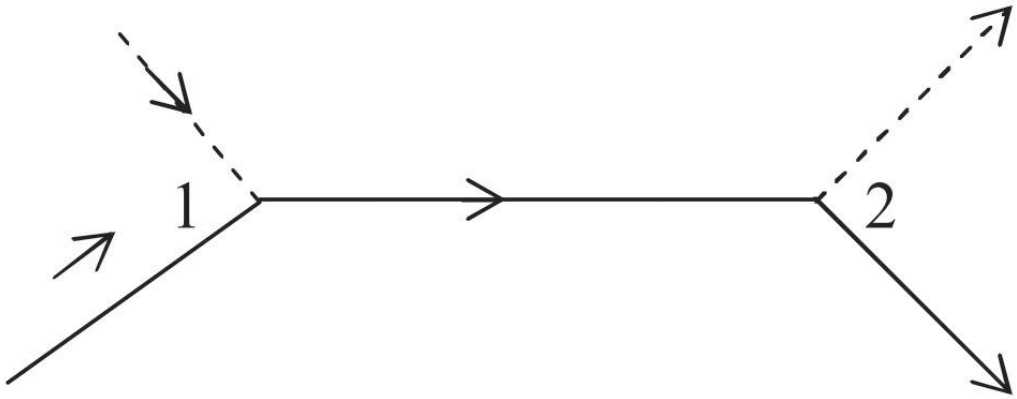


Figure 7.31.

In the initial state there is an electron and a photon. As they propagate, they meet at point 1. As a result of the encounter, an electron absorbs a photon. The resulting virtual electron emits a new, finite photon at point 2, which is emitted in one direction and propagates itself in the other direction. The result of the scattering process considered is a change in the initial directions of propagation of both the photon and the electron. The diagram consists of two vertex parts, the emission of the virtual electron at point 1 and its absorption at point 2.

Figure 7.32 shows a diagram of the interaction of two electrons. In the initial state there are two electrons that propagate in the same direction. If the first electron at point 1 emits a virtual photon in the direction of the second electron, then it, at point 2, is absorbed by the second electron. As a result, further directions of propagation of electrons diverge, i.e. electrons are repelled. The diagram consists of two vertex parts, the emission of a virtual photon at point 1 and its absorption at point 2.

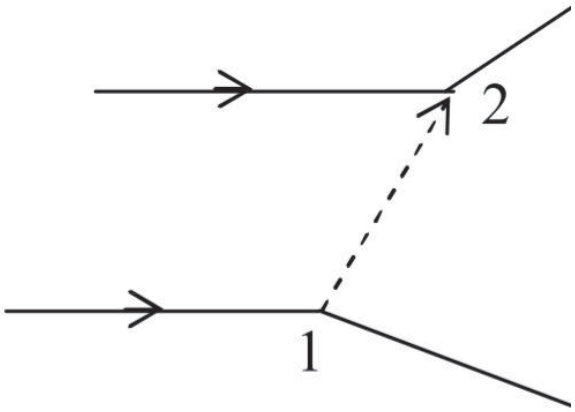


Figure 7.32.

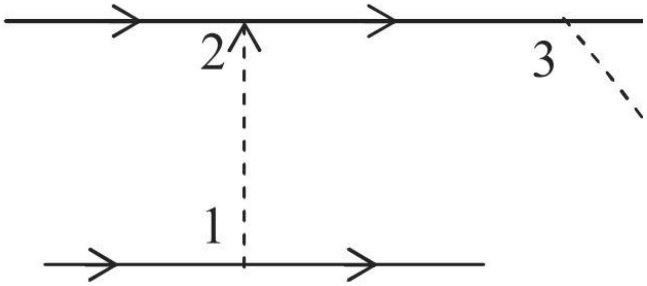


Figure 7.33.

Figure 7.33 shows a third-order diagram of photon braking radiation in the interaction of two electrons. Two electrons are present in the initial state. At point 1, the first electron, meeting the second electron, is inhibited and emits a virtual photon, which is absorbed by the second electron at point 2. Having been excited, the second electron radiates a non-virtual photon at point 3, resulting in braking radiation. The diagram consists of three head parts, the emission of a virtual photon at point 1, its absorption at point 2, and the emission of a photon at point 3.

An analysis of Feynman diagrams showed that the contribution of each vertex to the process amplitude is proportional to the interaction constant. The interaction constant is determined by the square of the absolute value of the wave functions of the interacting particles (process

probability). Mathematically, this is expressed by the fact that in the relation for the amplitude of the process, the coupling constant enters as a power with an index equal to the number of vertex parts (vertices) of the diagram.

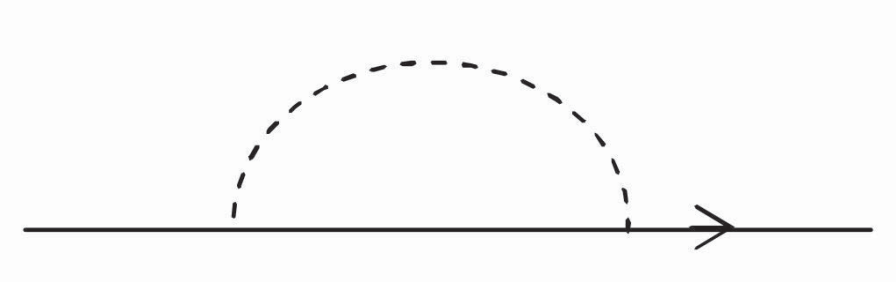


Figure 7.34

In addition to the interaction processes of two or more particles, self-interaction cases are also possible, when a virtual quantum of the corresponding field is emitted and absorbed by the same particle. A diagram of the **self-interaction** process is shown in Figure 7.34. The self-interaction leads to the formation of the particle's own field. It is created by the emission of a virtual quantum particle and its subsequent absorption by the same particle. It is not difficult to understand that the presence of self-interaction leads to an increase in the mass of the particle. This is explained by the fact that the field generated by the particle has some energy. According to the theory of relativity, the **field mass** m_F corresponding to this energy is added to the **non-field**, the so-called **priming mass** m_0 of the particle. Since the particle is assumed to be point-like, the field mass increases sharply as it approaches the particle and in the limit tends to infinity. For this reason, the self-interaction diagram is said to **diverge**. A similar divergence appears in the case of **polarization** of the physical vacuum arising in the presence of material particles. It has already been said above (see Section 7.6.1) that the physical vacuum is a virtual medium populated by virtual particles, with the help of which the processes described by the uncertainty relations are accounted for. This means that in the physical vacuum virtual particle-antiparticle pairs are continuously and randomly born and disappear (annihilate). The energy

of virtual field quanta is absorbed at pair birth, and this energy is emitted during annihilation in the form of virtual field quanta. Figure 7.35 shows a diagram of this process for the electromagnetic field. Under the action of a virtual photon emitted by the electrostatic field located at the point marked with an asterisk, a virtual electron-positron pair is born in the physical vacuum at point 1. Normally this pair annihilates at point 2. If an electron is brought to point 3 of the field, it interacts with the virtual pair. This interaction manifests itself in the fact that the electron at point 3 attracts the virtual positron of the pair and repels its virtual electron. As a result, the electron is surrounded by a layer (cloud) of virtual positrons, reducing (shielding) its charge. This phenomenon was called vacuum polarization. Since the electron is assumed to be point-like, the positron cloud includes an infinite number of positive charges, with the result that the effective charge of the electron would have to tend toward infinity. If the source of the primary virtual photon is another electron, then the process takes on the character of an exchange of electrons by virtual photons, and if one electron emits a virtual photon, then the other absorbs it. Due to the recoil in the exchange of photons, the electrons diverge, i.e. repel. When an electron exchanges photons not with an electron but with a positron (positive charge), the latter emits a photon in the opposite direction, and the interacting particles converge due to recoil, that is, are attracted.

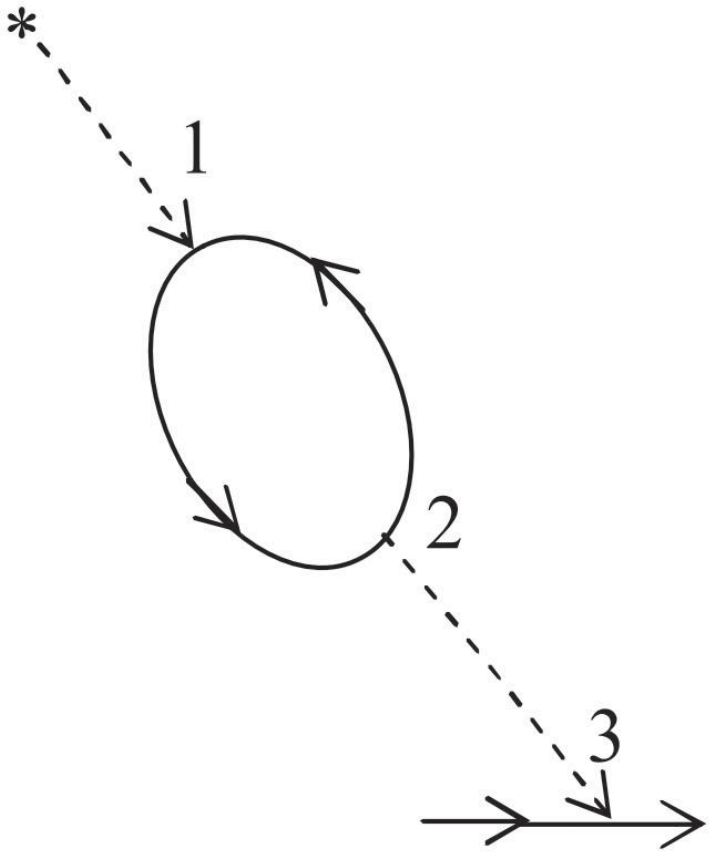


Figure 7.35.

The divergences, as mentioned above, are the result of the approximation underlying perturbation theory, according to which elementary particles are assumed to be point-like. This approximation is justified when considering processes that occur far enough away from the particles. This circumstance makes it possible to exclude divergences using the **renormalization method**. The specified method consists in using experimental values for the effective values of particle mass and charge. This makes it possible to isolate the finite observable parts of the indicated quantities. However, the renormalization method proved effective only in quantum electrodynamics, that is, in the region of relatively low energies. For this reason, an intensive search for new methods of describing microprocesses at all levels, i.e. in the field of high and ultrahigh energies, is currently under way.

7.6.8. Fundamentals of Quantum Gravity

The concept of gravity first entered physics with Newton's law of universal gravitation. However, this law only generalized the observations available at the time and did not explain the nature of the mutual attraction of bodies.

The first successful attempt to explain gravity was made by Einstein, who showed that gravity is a consequence of curvature of space-time by massive matter distributed in it and the field arising from gravitational redshift (loss of energy of photons as they move away from the massive body) which slows the rate of time flow. In this case, gravitation is perceived as an ordinary force that causes local acceleration of bodies, which, however, does not depend on their mass. This allows a quantitative description of gravitation, as in the case of other forces, using notions of the force field. In this sense, taking into account the theory of proximity following from the special theory of relativity, we can assume that the source of the gravitational field is the mass of bodies, and its carrier is gravitational waves.

The general theory of relativity revolutionized the idea of the mutual gravitation of bodies. However, in two of its most important provisions, it has remained in the position of classical mechanics.

First, it proceeds from the notion of the **continuity** of space-time. If we take into account that space-time is not an independent object, but only a reflection of properties of an evolving material totality, then quantization following from quantum theory should extend to space-time as well. Indeed, quantization follows already from the grand unification theory as much as from several other aspects of quantum theory. In the macrocosm, the discreteness of space-time does not manifest itself and can be ignored here. However, in the microcosm its accounting becomes mandatory.

Secondly, the general theory of relativity has remained in the position of the so-called **principle of determinism**. This principle was introduced into science by Pierre Laplace. Relying on Newton's laws of mechanics, Laplace assumed that there must be an exhaustive set of laws of Nature that would predict absolutely everything that has happened, is happening, and will happen in the Universe. For this you must know its state at this point in time. The celestial mechanics created by Laplace was a brilliant confirmation of the principle of determinism. For example, solar and lunar eclipses, great planetary confrontations and other events in the Universe, calculated according to Newton's laws of mechanics, were predicted by Laplace for many hundreds of years ahead with great accuracy.

The principle of determinism has long been regarded as the fundamental law of Nature. However, with the development of quantum theory in the first quarter of the 20th century, which emerged in connection with the penetration of science into the depths of matter, it became clear that microprocesses are described by uncertainty relations and are not subject to the principle of determinism. As a result, gravitational processes in the field of ultrahigh energies, i.e. in extremely small space-time scales do not obey neither general relativity nor quantum theory and at the same time in a number of aspects correspond to both theories. This means that for an objective description of gravitational processes in the field of ultrahigh energies, it is necessary to create a third theory that combines the first two, but does not contain the above contradictions. Such a theory was called **quantum gravity**.

The idea of unification is not far-fetched. It lies in the fact that in the region of high energies the difference between objects observed at relatively low energies is lost. So, for example, in the case of fast spinning wheel in a game of roulette, the ball, spinning non-stop, over time does not change its behavior. However, when the wheel slows down and the energy of the ball decreases, it falls through one of the thirty-seven grooves of the wheel. Since we cannot predict in advance which groove

the ball will fall into, we would have to conclude, based on the principle of determinism, that not one but 37 different types of balls are involved in the game. In other words, the various properties of the various elementary particles observed in the region of relatively low energies are leveled at the transition to higher energies. As a result, seemingly different particles in the low-energy region are combined into closely related groups (multiplets), the members of which can be regarded as different states of the same particle.

The same applies to fields, which, from the point of view of quantum theory, are also sets of discrete elementary particles, which are exchanged by matter particles during interaction. Thus, for example, electromagnetic interactions, on the one hand, and weak interactions, on the other, in the region of sufficiently high energies, on the order of 102 -103 GeV, lose their identity. This is because the quanta of these fields, photons and intermediate vector bosons, at these energies form a single multiplet of almost identical particles. As a result, the electromagnetic field and the field of weak interactions combine to form a single field of electrically weak interactions.

At present there is no complete quantum theory of gravitation yet. There are only a number of approaches to it. Let us consider some of them, the most logical and most closely reasoned.

7.6.8.1. Supersymmetry

Supersymmetry, as stated above, relates fields whose quanta have zero or integer spin and belong, respectively, to bosons and fields whose quanta have half-integer values and belong to fermions. Fields in supersymmetry transformations form supermultiplets describing particles with the same mass but different spins. At zero mass, the supermultiplet includes particles with spins I, I+1/2, and at non-zero mass - with spins I-1/2, I, I+1/2.

Quantum field theory distinguishes between **spinor**, **scalar**, and **vector** fields.

A **spinor field** is a physical field that is described by the function $\psi(\vec{r}\,,t)$, which is a **spinor** at each point, i.e. it consists of two components that mutually transform into each other as the coordinate system rotates by a certain angle. The notion of a spinor was introduced in 1927 by Wolfgang Pauli, who generalized the Schrödinger equation by taking into account the spin of a charged particle equal to 1/2 (the Pauli equation). Such a particle can be in two different spin states with spin projections equal to +1/2 and -1/2. Accordingly

The wave function describing this particle is two-component, and is written in the form of a matrix-column

$$\Psi(\vec{r},t) = \begin{vmatrix} \Psi_1(\vec{r},t) \\ \Psi_2(\vec{r},t) \end{vmatrix} \qquad (7.180)$$

This function is called a **spinor**. For the first projection of spin +1/2 $\psi = \psi 1$ and $\psi 2 = 0$, and vice versa. Quanta of the spinor field are all fermions with spins equal to 1/2, namely electrons, muons, neutrinos, quarks, and their antiparticles.

A **scalar field** is a physical field that is described by a scalar function φ(r,t) that does not change as the coordinate system rotates. Such a particle is a boson with spin equal to zero. Even functions are called **scalar** and odd functions are called **pseudoscalar**.

A **vector field** is a physical field that is described by a vector function. The quantum of the vector field is a boson particle with spin equal to 1.

These functions can be converted to each other. Let in the simple case the supersymmetry transformation transforms the spinor function ψ to the scalar function φ by an angle ε. For small angles we can obviously write that

$$\Psi(r) = \psi(r) + \varepsilon\varphi(r) \qquad (7.181)$$

The equation (7.181) is similar to the relation describing the shift and transition from one point to another in space-time. In this sense, supersymmetry is local.

Supersymmetry means insensitivity to the choice of different components of the supermultiplet. Supersymmetry contains all components necessary to describe weak and electromagnetic interactions - spinor particles (leptons and quarks), vector particles (photons, intermediate vector bosons), and scalar particles. The supersymmetry condition establishes relations between masses of the specified particles and interaction constants. In real conditions supersymmetry must be broken, because in Nature there are no quanta of fields, which are fermions, as well as no spinor particles with zero mass. Hence, by the way, it follows that the neutrino cannot be a massless particle.

A supersymmetric generalization of gravity theory is called **supergravity**. It includes transformations obeying the relation (7.181). Supersymmetric invariance, by analogy with gauge invariance, introducing a photon as a gauge massless particle with spin 1, leads to the necessity of introducing gauge massless particles with spins equal to, 3/2 and 2. The first of them is called **gravitino** (by analogy with neutrino), and the second is identified with **graviton**, which according to quantum field theory is a carrier of gravitational interactions. The local generalization of the **extended supersymmetry**, covering both space-time and internal degrees of freedom, is called extended supergravity. Extended multiplets contain, in this case, in addition to the above particles, also particles with spins 1, 1/2, 0. In other words, extended supersymmetry combines all four fundamental interactions. At the same time it is not possible to place all known elementary particles in the above scheme, which does not allow to put it as a basis for quantum gravity theory yet.

7.6.8.2. The Cosmological Hypothesis

Another approach to the quantum theory of gravity was proposed by one of the most famous among modern physicists theorist English cosmologist Stephen Hawking. The hypothesis he proposed is closely connected with cosmological concepts and, in particular, with the modern theory of the evolution of the Universe, the starting point of which is the cosmological **singularity**, which he also introduced in 1970.

Back in 1965, another English physicist Roger Penrose showed that under the action of the so-called **gravitational collapse** a star under certain conditions is indomitable compressed, so that its volume, mass and energy are reduced to zero, and the density and curvature of the surrounding space-time are increased to infinity. As a result, a point superdense formation appears in space-time. This statement is called the Penrose theorem. The phenomenon of gravitational collapse is quite common in the Universe. It occurs whenever, due to fluctuations or explosions of so-called supernovae, which eject the internal matter of a star into interstellar space, a condensation of matter is formed in a particular region. Under the influence of increased gravity, the compaction, and, consequently, its mass, increases avalanche-like. This growth is compensated by an increase in the electromagnetic repulsion of the particles of matter as they approach each other, as well as by their pushing away in accordance with the Pauli exclusion principle. Calculations show that if the mass of the arising cosmic formation does not exceed 1.25 solar masses, then further growth of its mass stops and a compact stable star is formed. If the mass of the star exceeds 3 solar masses, gravity overcomes the star's internal resistance, and it begins to shrink inexorably, according to Penrose's theorem. Analyzing this theorem, Hawking came to the conclusion that when time reverses in the model of the Universe proposed in his time by the Russian scientist from St. Petersburg Alexander Friedmann (see 3.9), its initial state must also be a singularity. Meanwhile, when formulating Penrose theorem and Hawking hypothesis quantum effects were not taken into account, which

cannot be absent in singularity, which is a quantum object due to its microscopicity.

Let the mass of the star contracting into singularity be equal to the value *M*. If a body of mass *m* appears near the star, at a distance *r*, then under the action of gravity it acquires energy equal to

$$\mathcal{E}_{\mathrm{p}} = G\frac{Mm}{r}, \tag{7.182}$$

On the other hand, accelerating in the gravitational field to the speed of light, the body receives kinetic energy, which tries to bounce it away from the star, equal to

$$\mathcal{E}_{\mathrm{k}} = \frac{mc^2}{2}. \tag{7.183}$$

The mass of the star M is determined from the uncertainty relation for momentum, namely

$$M_{cr} \approx \frac{\hbar}{2}. \tag{7.184}$$

As long as the kinetic energy of the body is less than the gravitational energy, it can approach and move away from the singularity. From the moment the gravitational energy becomes greater than the value (7.183), the body starts inevitably approaching the singularity. The distance r_{pl}, at which the equality of energies occurs, is called the **Planck radius**.

From equations (7.182), (7.183) and (7.184) it follows that the Planck radius

$$r_{pl} = \sqrt{\frac{G\hbar}{c^3}} = 1.6 \cdot 10^{-35}\ m \tag{7.185}$$

The Planck radius defines the boundary of the singularity **horizon**, also called the **Schwarzschild sphere**, and the singularity together with its horizon form the so-called **black hole**. This name arose

because any object, including a quantum of any field, such as a photon crossing the horizon line, falls into a black hole and cannot come out of it anymore. Since this also applies to a set of photons, i.e., rays of light, a black hole cannot be seen. Not only cosmic bodies evolve into black holes, but also microparticles. A microparticle approaching a particle interacting with it at a distance of the Planck radius increases its speed to the speed of light. As a result, it sharply increases its mass (see Section 3.5.1) and gravitational collapse occurs. One of the particles turns into a black hole, and the other particle falls into it. This means that the size of particles, and therefore the size of the element of space cannot exceed the value of the Planck radius. This minimum possible size of the spatial cell is called its quantum. A **quantum of space** corresponds to a **quantum of time**, which is apparently equal to $0{,}5 \cdot 10^{-43}$ s.

From the idea of space-time quantization it follows that the singularity cannot shrink into a point and is finite, and from the uncertainty relation for energy, taking into account space-time quantization, we obtain for the density of matter in the singularity

$$\rho \approx 10\ 96\ \text{kg/m}^3.$$

The idea of quantization of space-time can also be considered as a theoretical justification of the postulate of the theory of relativity about the finiteness of the speed of propagation (transfer) of all kinds of interactions in Nature.

Based on quantum concepts, singularity cannot be regarded as a purely material object. In it, as in other quantum objects, processes have a distinctly non-deterministic nature. In other words, it does not fulfill the laws of conservation of energy-momentum, causality, and other so-called objective laws of Nature, which are attributive to material objects. Wishing, however, to keep traditionally familiar approaches, it is possible, as it is accepted in quantum field theory, to consider singularity as a point, purely material object with very high, but still finite density,

non-zero mass and energy. These quantities are calculated according to the uncertainty relations and notions of space-time quantization. The non-material, quantum character of singularity is taken into account in the same way as in quantum theory of interactions, by introducing the notion of physical vacuum (see item 7.6.1).

The singularity concept was met with great distrust and was severely attacked, especially by Soviet scientists. However, it has now found almost universal acceptance. Meanwhile, it cannot be claimed that all the problems associated with the singularity notion have been finally solved. Studies of the various models of the Universe at its initial moment and of the ways in which it evolved to its present state indicate that its present state may have arisen from a large number of different initial configurations. In order to predict what this beginning must have been, we cannot limit ourselves to relativity theory alone, which, as we have already pointed out above, does not take into account the quantum effects that inevitably arise at the singularity. On the other hand, modern quantum theory does not take into account the influence of powerful gravity, which takes place in the singularity. Moreover, there is every reason to suppose that at transition to minuscule space-time intervals of singularity in the region of huge energies acting there the laws of modern quantum theory also become invalid. This means that in order to correctly judge what the beginning of the Universe really was, it is necessary to supplement field theory with **quantum gravity theory**. This theory should unite quantum field theory and relativity theory. Among the various hypotheses underlying this future theory, the Hawking hypothesis stands out, which, by the way, shows how much the current ideas about the initial state of the Universe can change within the framework of this new theory.

Let us consider the main points of Hawking's hypothesis. We emphasize, first of all, that according to Hawking, the theory of quantum gravity should organically include the main provisions of existing theories - relativity theory and quantum physics. This means that the new

theory must necessarily include the provision of Einstein's general theory of relativity that the gravitational field appears as a curved space-time, in which motion by inertia occurs not along straight lines, but along curved geodesic lines. It should also include the Feynman method of quantum theory, according to which the motion of a particle is not characterized by a single trajectory, as in classical mechanics, but is considered as movements along all possible trajectories of space-time, which are equivalent to some spherical wave according to wave-particle duality. In this case the position of the particle at any given time is described not by the coordinates, but by the wavelength and phase. The latter means that the probability of a particle passing through a given point in space-time is determined by summing all the waves corresponding to each possible trajectory passing through that point.

In the classical theory of gravity, that is, the general theory of relativity, only two types of behavior of the Universe are possible - either it existed forever, or its beginning was a singularity. This alternative is a consequence of the fact that the Universe is considered here in actual time, in which the spatial coordinate axes and the time axis are in different spheres (see Section 3.6). The axis of time always lies inside the cone of light, which defines all possible events of the past, present, and future, and the axis of space lies outside of it. In simplified terms, this can be explained as follows. Let the light impulse is emitted at some point in time from some point in space, for example a point on the surface of the Sun. As time passes, it spreads out in all possible directions at the same speed and becomes a sphere of light. Let's represent this imaginary as a space-time diagram, with the axis of time pointing vertically upward. Let the light sphere collides with some cosmic body, the surface of which cuts a light cone from it with the apex at the point of emission of the light impulse. A body surface perpendicular to the time axis is a two-dimensional spatial coordinate surface whose axes are outside the light cone. Symmetrically to this light cone there is the same light cone, the time axis of which is directed in the opposite direction. One of these cones

contains the set of all events that can affect an event at a given point. This cone is called the cone of the past. The second cone, the cone of the future, contains the set of all events, which in principle can be affected by what is happening at this point now (see Figure 3.4). The outer region in relation to the light cones is characterized by the fact that the events belonging to it can neither affect the events at that point nor be influenced by the events occurring at that point. Thus, although space and time form a single continuum, they are unequal and formally oppose each other as imaginary and real categories.

This circumstance greatly complicates the summation problem for all particle trajectories. There is, however, a simple method that makes this task easier. It consists in the fact that the waves that form those trajectories of particles, whose motion occurs not in real but in imaginary time, are added up. In this case, all coordinate axes, including time, become equal. Hawking's hypothesis is to combine Feynman's summation along the trajectories of quantum theory with the notions of the theory of relativity about gravitationally curved space-time. This essentially means moving on to viewing the Universe from the unified perspective of quantum gravity. The trajectories of the particles are curved as if gravity were acting on these particles. The analogue of all possible trajectories of a particle becomes all curved space-time, which represents the history of the Universe, and summation by trajectories is replaced by summation by histories. This approach allows us to add a third type of its possible behavior to the alternative development of the Universe discussed above. For this purpose it is enough to pass to consideration of the Universe in imaginary time, in which all axes of four-dimensional space-time are equal, and finite space-time is boundless, i.e. devoid of singularity forming its boundary. The **spherical** surface of the Earth can serve as a visual model of such space-time, but not in four-dimensional, but in three-dimensional consideration, if we consider the axis of time as its vertical axis, replacing, however, the idea of real time by imaginary time. Taking the concept of a finite but boundless Universe

as the basis, its history can be conventionally depicted in the form of the Earth's surface according to the above model. Then the distance to the North Pole corresponds to imaginary time, and the dimensions of the circles, all points of which are equidistant from the pole, correspond to the spatial dimensions of the universe at a given time. In this case, the beginning of the Universe still corresponds to a point. This point is at the North Pole. However, unlike a singularity, it is ordinary and physically no different from any other point of the spherical surface under consideration, which represents a point of the Universe at any moment of its evolution. In this sense, the North Pole point has nothing to do with the singularity, in which all the laws of Nature are violated. The expansion of the Universe in this case is equivalent to the movement from the North Pole to the equator, and the movement from the equator to the South Pole corresponds to its contraction.

Hawking's hypothesis boils down to the assumption that: "...so-called imaginary time is really the real time, and that what we call real time is just a figment of our imaginations...".

We emphasize, first of all, that there is nothing irrational in Hawking's hypothesis. It is known that our subjective feelings are not only far from reality, but sometimes have a very distant relationship to it. The world of images, colors, sounds, smells, tastes, thermal sensations, etc. is not reality itself, but only a reflection of some very distant from reality orders, about the essence of which we still have approximate, very distant from reality ideas. Inverted images are perceived by us as straight and imaginary images as real. Therefore, it would not be surprising if we suddenly learned that the time we experience is in fact imaginary. In that case, though, we would have to reconsider all our ideas about the real world, and with them all science. After all, not only time would be imaginary, but many other quantities, such as speed, momentum, power, current, components of the electromagnetic field, etc. Many purely positive quantities, such as kinetic energy, would be negative. The signs of many other quantities would be reversed. It is hard to believe that

Nature would go to these complications and introduce such confusion into our ideas.

Nevertheless, the degree of reliability of Hawking's hypothesis, as well as any other hypothesis, can only be judged by the results of observation and experiment. One of the most fundamental predictions of this hypothesis boils down to the assertion that the finite Universe is boundless. It is easy to show that a direct consequence of this assumption is really observed in practice isotropy of intensity of the **relic** radiation. Another consequence of Hawking's hypothesis is the inhomogeneity of the density of the Universe in its initial state, which follows from quantum theory and, in particular, from the uncertainty relations.

It should be kept in mind, however, that these facts are not solely the consequences of the boundlessness of the Universe, but can be derived from other considerations unrelated to Hawking's hypothesis. Thus, we can only speak of a possible confirmation, not a proof of Hawking's hypothesis. It is noteworthy, however, that in spite of this, assertions about the boundlessness of the finite Universe are widely accepted in the scientific literature, although, strangely enough, they are combined with a simultaneous recognition of the singularity as a beginning. It follows that Hawking's hypothesis has the right to exist, but it cannot be considered proven.

7.6.8.3. Superstring Theory

In 1968, two young physicists from CERN (European Center for Nuclear Research), Gabriele Veneziano and Mahiko Suzuki, independently of each other, noticed that the scattering amplitude of high-energy pions can be expressed by a formula using the so-called Euler beta function (see Appendix 2). Two years later, a number of physicists, including Leonard Susskind, Holger Nielsen, and others, who were dealing with the same problem, discovered that the above formula could be derived differently. To do this, it was necessary to assume that interacting pions are connected by an infinitely thin, taut string, which is

described as a quantum-mechanical object. As a consequence, they had the idea that the string in question could be regarded as a kind of model of an elementary particle. Generalizing this idea, they put forward the hypothesis that the interaction of elementary particles can be compared with the connection and disconnection of these strings. Standing waves occur in such a taut string. The number of half-waves stacked on the length of the string determines the mode of the oscillating string. Each mode corresponds to a certain type of particle. Note that for physicists working in the field of quantum theory, there was nothing unusual in this hypothesis. The identity of particles and waves has been known for a long time, since the formation of the idea of wave-particle duality. Subsequently, the similarity between elementary particles and waves became one of the cornerstones of quantum mechanics and formed the basis of Schrödinger's wave equation. The string in this sense can obviously also be regarded as a kind of wave-particle formation. This is how the string theory emerged. Its appeal initially lay in the fact that it excluded point objects from the theory (a string has, although small, but finite length) and, with them, divergences. At the same time it had a lot of disadvantages. First, from the very beginning, a large number of versions of string theory, independent of each other, appeared, creating serious confusion. For example, there was a separate theory for particles with integer spin, bosons, a separate theory for fermions, etc. The theory led to anomalies of various kinds and, in particular, to a violation of the law of conservation of energy. The original string theory did not allow to isolate from an infinite sequence of modes those which corresponded to a finite and relatively small number of real particles. Back in 1970, the American physicist Claud Lovelace showed that the Veneziano model, which formed the basis of string theory, is based on a **26-dimensional** space-time continuum. By introducing spin into string theory, John Henry Schwarz concluded that in this case the number of dimensions is reduced, and the theory is realized in **10-dimensional** space or in 11-dimensional space-time. It is known that any physical theory is true for any number of measurements. Unexpectedly it turned out that the string theory was the

only one of its kind, which was performed only for a given dimensionality, which in addition exceeds the number of dimensions of the real world that we are accustomed to.

All this at first undermined physicists' confidence in the new theory. However, by now the situation has changed significantly. First, restrictions were introduced into the theory, which allowed us to exclude from consideration those modes that do not correspond to real particles. The new theory was called the superstring theory. In 1984 John Henry Schwarz and Michael Green proved that the anomalies found in the superstring theory mutually compensate each other, just as in quantum theory, the energy disappearing at the birth of particles and arising during their annihilation is compensated, so that on average the law of conservation of energy is not violated. In 1995, the American physicist Edward Witten showed that the different versions of string theory that existed before were in fact only different expressions of the same superstring theory. This theory is called the **M-theory**. It is, as the same Witten showed, 11-dimensional. The greatest success of superstring theory was the discovery among the solutions of string equations of a closed string, which corresponds to a particle with zero rest mass and spin equal to 2. This came as a surprise, because according to quantum field theory a hypothetical massless particle with spin equal to 2, called graviton, is known to be a quantum of the gravitational field and corresponds to the gravitational waves predicted by the general theory of relativity. This led to the assumption that the mysterious particle following from the superstring theory corresponds to the graviton, which could in principle allow to combine quantum theory and Einstein's theory of gravitation and build the long conceived, but elusive within quantum field theory, general physical theory of quantum gravity.

Since then, interest in superstring theory has increased again. This was also facilitated by Witten and Townsend's original interpretation of the multidimensional nature of superstring theory. According to this interpretation, the string itself is one-dimensional or, given time, two-

dimensional. In string theory, a string is said to sweep a two-dimensional surface over time, called a membrane or two-dimensional brane. The freedom of string movement is limited, however, by the complementary spatial manifold, which corresponds to 3-branes, 4-branes, etc. up to 9 branes. From our point of view this may mean that branes, impose on each elementary particle certain bonds, limiting the mode of vibration of the corresponding string of a given length, ordering its internal structure and, thus, giving the particle a certain individuality. The restriction of freedom of movement of the string on the brane side occurs only if the string is not closed, that is, if its ends are free and secured in the respective branes. From the superstring theory, in particular, it follows that all elementary particles, except gravitons, correspond to strings with free ends. The ends of the strings can be attached to the same brane or to two different branes. It turned out that all known particles, except the graviton, whose string has no free ends and, therefore, is not fixed, correspond to strings, both ends of which are fixed on the 3-brane. It corresponds to 3-dimensional space. This means that neither they, nor all their formations, can leave it and go to other branes.

We, therefore, are in the 3-brane not because there are no other branes in the world, but because we are trapped in it and cannot leave it. Another thing is gravitons, they can in principle leave our three-dimensional world. This, however, is not happening.

There are two versions to explain this phenomenon. The first of them believes that gravitons remain within the 3-brane simply because they are attracted by it. The other one, which is more reasonable, proceeds from the fact that all dimensions, except for three corresponding to the real world, are rolled up in small space and appear only at very small distances, about 10^{-35} m (according to M-theory, at a distance of 10^{-18} m).

This hypothesis can in principle be easily tested experimentally. The point is that interaction forces must always be inversely proportional to distance to the power of (n - 1), where n is the dimensionality of space.

In three-dimensional space, for example, they are inversely proportional to the square of the distance (the law of universal gravitation, Coulomb's and Ampere's laws, etc.). Experimentally, the specified gravity dependence has been confirmed up to distances of 0.1 mm so far. What is the dependence of gravitational interactions on distance at very small intranuclear distances cannot yet be verified, although experimental physics is already close to solving this problem. In addition, the question arises why only 4 of the 11 dimensions have unfolded, including 3 spatial and time. It is interesting to note that if the dimensionality of real space were less than or greater than 3, then, as calculations show, the functioning of the Universe would be impossible.

Analyzing the current state of theoretical physics, Professor of the University of California Joseph Polchinski notes that the superstring theory opened the way to the construction of quantum gravity theory and a unified physical theory. From his point of view, it united in a single mathematical structure all four types of fundamental interactions, helped to solve many paradoxes that arose in the quantum theory of black holes, and freed the theory from the problem of divergences.

It is assumed that with the help of hadron colliders it is possible to organize a test of string theory. However, no confirmation of these hypothetical theories has been obtained at the LHC so far.

7.6.8.4. Fundamentals of the Loop Theory of Quantum Gravity

The loop theory of gravitation eliminates the differences between the general theory of relativity and quantum theory by treating space-time as a quantum object, which initially has wave-particle duality and a discrete internal structure. According to the loop theory, it consists of the smallest indivisible volumes (cells) and the areas bounding them, which change by discrete jumps, like the frames of a movie. Possible values of volume and area are measured in units derived from the Planck length. It defines the scale at which the geometry of space can no longer be considered continuous and smooth, equal to 10^{-35} m. The smallest

possible area other than zero is approximately equal to the square of the Planck length or 10^{-70} m^2. The smallest possible volume other than zero is the Planck length cube or 10^{-105} m^3. Thus, according to this theory, each cubic meter of space contains approximately 10^{105} atoms of space volume. The quantum of volume is so small that there are more such quanta in a cubic meter than there are cubic meters in the entire visible Universe (10^{79}).

Quantum states of volume and area are depicted in the form of diagrams, graphs according to the following rules.

Let us assume for simplicity that the region of space is shaped like a cube. In the diagram it (Figure 7.36a) is represented by a node as a point representing the volume of a cube, with six lines coming out of it, each line representing one of the faces of the cube (Figure 7.36b). The number next to the node indicates the size of the volume, and the numbers next to the lines indicate the size of the area of the corresponding faces. Let us place a pyramid on top of the cube (Figure 7.36c). Polyhedrons have a common face, and it is depicted as two points (two volumes) connected by one of the lines (the face that connects the volumes, Figure 7.36d). Drawings of polyhedrons are discarded and only graphs are left.

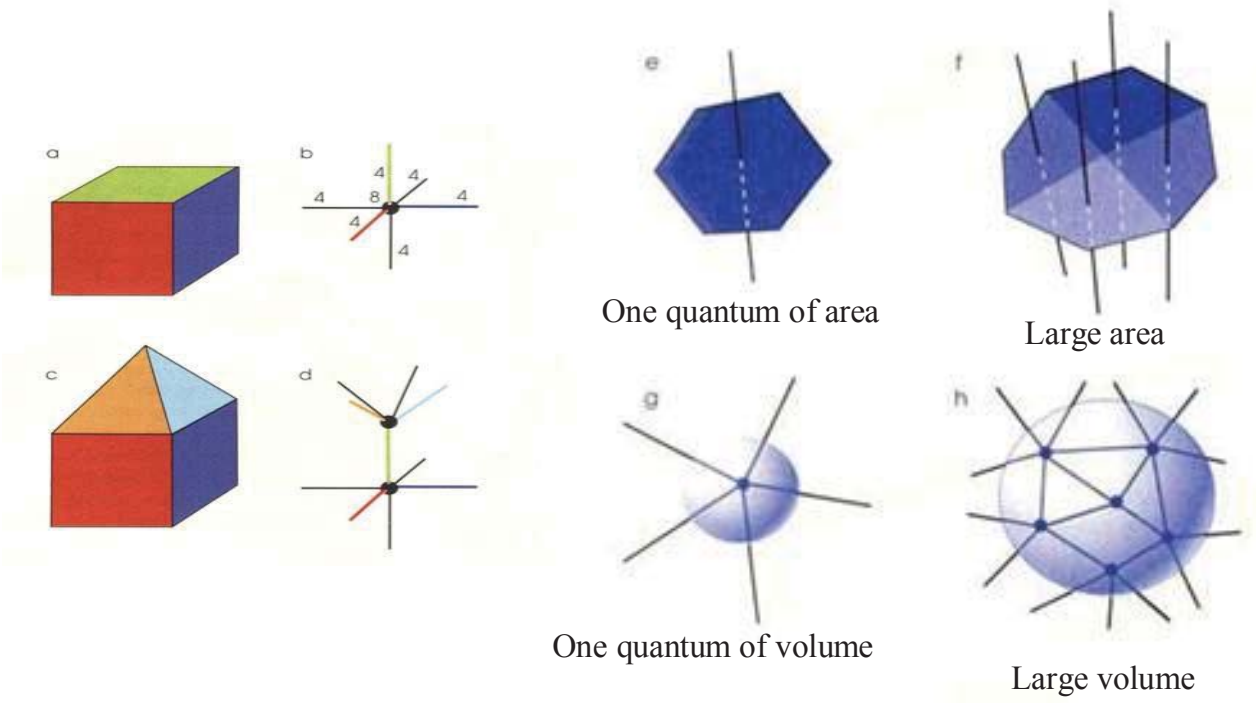


Figure 7.36.

Graphs in loop quantum theory are peculiar drawings of pieces of space and are called a spin network (Figure 7.36h). They describe not the configuration of the cell, of which we know nothing, but the quantum state, one of the main characteristics of which is spin. Each quantum state corresponds to one network cell (there are 10^{184} of them in space-time), and each network cell satisfying certain rules corresponds to a quantum state. It should be emphasized that the node representing the cell volume is not the capacity of the particle, but a potentially possible particle, which is indicated in the network by a certain label.

Since, in the general theory of relativity, space is inseparable from time, they should be considered together. Lines of the spin network in this case expand and become two-dimensional surfaces, and the nodes representing possible particles are stretched into lines. The nodes from which the lines come represent transitions from one state to another. Time, like space, is discrete. The minimum quantum of time is approximately 10^{-43} s, and the transition from one state to another is occured by leaps and bounds. Since each state corresponds to its own spin

network, space-time is represented by a set of spin networks, which is called a spin foam (Fig. 7.37).

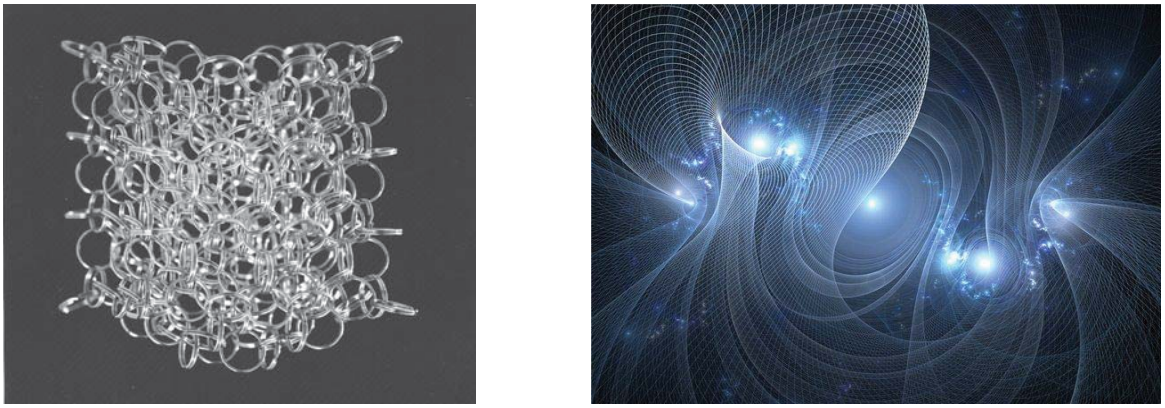


Figure 7.37

One of the advantages of the loop quantum theory of gravitation is the naturalness with which the **Standard Model** of elementary particle physics gets its explanation.

Harari-Shupe preon model was originally proposed. The term "preon" itself was used to refer to point subparticles included in the structure of fermions with half-integer spin (leptons and quarks). It was built on the same principle as the quark model.

However, the use of point particles leads to divergences. Therefore, Sundance Bilson-Thompsonn proposed a model based on the more general theory of twisted ribbons (braids), in which point Harari preons (rishons) were transformed into extended ribbon-like objects, ribbons. Potentially, this explained the emergence of the color charge of quarks, eliminating divergences and mass paradoxes. This model leads to an understanding of electric charge as a topological entity arising from ribbon twisting.

This theory assumes that excited states of space-time itself can play the role of preons, leading to the standard model of quantum gravity theory.

Bilson-Thompson and his co-authors suggested that the theory of loop quantum gravity could automatically unite all four fundamental interactions. Using ribbons, represented in the form of weaves of fibrous space-time, it was possible to build a successful model of the first generation of quarks and leptons with a more or less correct reproduction of their charges and parities.

In this model, it is assumed that the electric charge and color charges, as well as the parity of particles belonging to generations of higher rank, should be obtained in exactly the same way as for the first generation particles. The use of quantum computing methods made it possible to show that these kinds of particles are stable and do not disintegrate under the influence of quantum fluctuations.

Ribbon structures in the Bilson-Thompson model are represented as entities composed of the same matter as space-time itself. The question of how the **Higgs boson** can be obtained with the help of the loop theory is still open.

It is implied that properties of particles (their masses, energies and spins) may correspond to properties of loops - basic objects of the theory of loop quantum gravity.

Some other Standard Model particles, such as photons, gluons, and gravitons, can also be reproduced. The mass of the particles is determined in proportion to their internal structure, i.e. the twist of the braids. For example, in the Bilson-Thompson model, the structure of a photon having zero mass corresponds to three untwisted ribbons.

What is important for the Bilson-Thompson approach is that in his preon model elementary particles, such as the electron, are described in terms of wave functions.

Following the principle of quantum superposition and the principle of additionality, an attempt was made to combine the loop theory of gravitation and string theory.

Loop quantum gravity attaches less importance to matter present in space-time, and concentrates more on the properties of space-time itself. In this theory, space-time is a network, a material totality itself, and ribbons are like strings. Instead of oscillating, they twist, warping space-time accordingly. The smooth background of Einstein's theory of gravity is replaced by knots and links, which are given quantum properties. In this way, space is made up of individual pieces. Theory is mostly concerned with the study of these pieces.

This approach has long been considered incompatible with string theory. Indeed, their differences are obvious and profound. Loop theory studies pieces of space-time, while string theory investigates the behavior of objects in space-time external to them. These areas also share technical challenges. According to string theory, there are 11 dimensions in space-time, and according to loop theory there are the usual 4 dimensions. String theory suggests the existence of supersymmetry, in which all particles have undetectable partners. Supersymmetry, however, is not peculiar to loop theory.

It is true that recently the opinion of physicists has changed. New theoretical discoveries have revealed possible similarities between loop theory and string theory. Attempts to solve some problems of loop theory led to the discovery of its first unexpected connection with string theory

The loop theory initially had a deviation from the special theory of relativity. In the special theory of relativity, the linear dimensions of an object are reduced only from the point of view of a relatively stationary observer. And in the loop theory, as the relative velocity of an object increases, it is compressed, similar to Lorentz contraction, in an absolute sense, affecting the size of pieces of space-time as well.

It turned out, however, that bringing the loop theory into conformity with the special theory of relativity inevitably entails the appearance of interactions similar to those present in string theory. In addition, a group of physicists at the University of Erlangen–Nuremberg (Germany) managed to include supersymmetry in the loop theory, which had previously been the sole territory of string theory.

7.6.9. Advances and problems of quantum field theory.

Quantum field theory is a unique, though quite natural step in the development of physics. Its importance for the development of all physical science can hardly be overestimated. For the first time in the history of science, it helped to explain the nature of interactions at different levels of the real-world systems that condition all the processes going on in it. It also for the first time came very close to creating a unified physical theory based on the structural elements "bricks" and physical fields of a single nature underlying the Universe. Quantum field theory, along with quantum theory of elementary particles, made it possible to approach the understanding of problems related to the evolution of the Universe. They explained the formation of the entire spectrum of chemical elements and matter in general, as well as biochemical evolution and the emergence of life. Quantum electrodynamics, created in relation to the interaction of electrons, positrons and photons and considering the interaction between electrically charged particles as a result of their exchange of virtual photons, became the foundation of quantum field theory. Electromagnetic field in quantum electrodynamics is manifested as a gauge field, the quantum and gauge particle of which is a photon. The ideas underlying quantum electrodynamics, whose conclusions proved to be in perfect agreement with experience, were later extended to other types of interactions, which also found experimental confirmation. In particular, this is how chromodynamics, which deals with the laws of strong

interactions, and the unified theory of electrically weak interactions, emerged.

It does not follow, however, that field theory and quantum theory as a whole do not have any difficulties and unsolved problems. Let us point out some of them.

1. The problem of divergence.

This problem is that in some cases the theory leads to infinite values of masses and charges of interacting particles. Such divergences appear, for example, in the calculation of the self-interaction of fields, in the processes of pair birth and annihilation, etc. It should be emphasized, however, that the problem of divergence does not arise in field theory itself, but in its currently accepted mathematical apparatus, perturbation theory and Feynman diagrams. These theories, as indicated, consider elementary particles as point formations. This approach ignores quantum-mechanical properties of elementary particles and space-time in the region of high energies, i.e. in the depths of material objects. For this reason, it is purely approximate and admissible only when considering processes occurring in the region of relatively low energies. To increase the degree of the conventionally accepted approximation of perturbation theory, the so-called renormalization method is used, which consists in using experimental values of computed divergent quantities and separating finite observable parts from infinite computed values. The renormalization method gives acceptable results only in the region of electromagnetic, weak, and electrically weak interactions, that is, in the region of still relatively not very high energies. In the region of higher energies, which are characteristic of strong interactions, it is not always possible to get rid of divergences.

2. The problem of non-identity of particles with the same spins.

In particular, the reason for the division of particles with spin 1/2 into quarks and leptons has not been fully understood. The above

assumption that the difference in the properties of these particles is a consequence of breaking the lepton-quark symmetry in the high energy region is hypothetical, since experimental verification of this hypothesis is impossible, at least at the present or close to it high energy levels, because of the impossibility of reaching these high energies.

3. The problem of free state of quarks and gluons.

The existence of quarks and gluons is hypothetical because of the fundamental impossibility of obtaining them in the free state and observing them by existing methods.

4. Problems related to the nature of physical constants, including charges. Quantum field theory, like other theories, has not solved these problems so far.

5. The problem of the imperfection of the experiment.

Until very recently, the experimental technique did not provide an opportunity for reliable experimental verification of results, hypotheses, and models derived from field theory. For example, for a long time it was impossible to verify the consequences derived from the theory of gauge symmetry and to determine, in particular, the nature of internal quantum numbers. Only in 1983 it became possible to detect and measure the masses of the components of intermediate vector bosons carrying weak interactions, which had long before been predicted by theory. This became possible after the discovery of the Higgs particle, which is responsible for the magnitude of elementary particle masses. However, a new problem has arisen in connection with the fact that it is now necessary to explain the nature and magnitude of the mass of the Higgs particle itself. Up to now it is not possible to detect hypothetical gravitons and gravitino, magnetic monopoles, etc., predicted by the theory of gauge symmetry and supersymmetry. The quantization of space-time, the instability of protons and their lifetime predicted by the theory also cannot be confirmed.

6. The problem of quantum gravity theory.

Einstein's general theory of relativity is classical in the sense that it ignores quantum effects occurring in the high energy region, i.e. in small space-time regions, such as the vicinity of black holes, the depths of matter, etc. This puts on the agenda the problem of creating a more general theory of quantum gravity, acting also in the high energy region, which would combine general relativity theory with quantum field theory into a single theory, incorporating them as special cases and limiting their action to the low energy region.

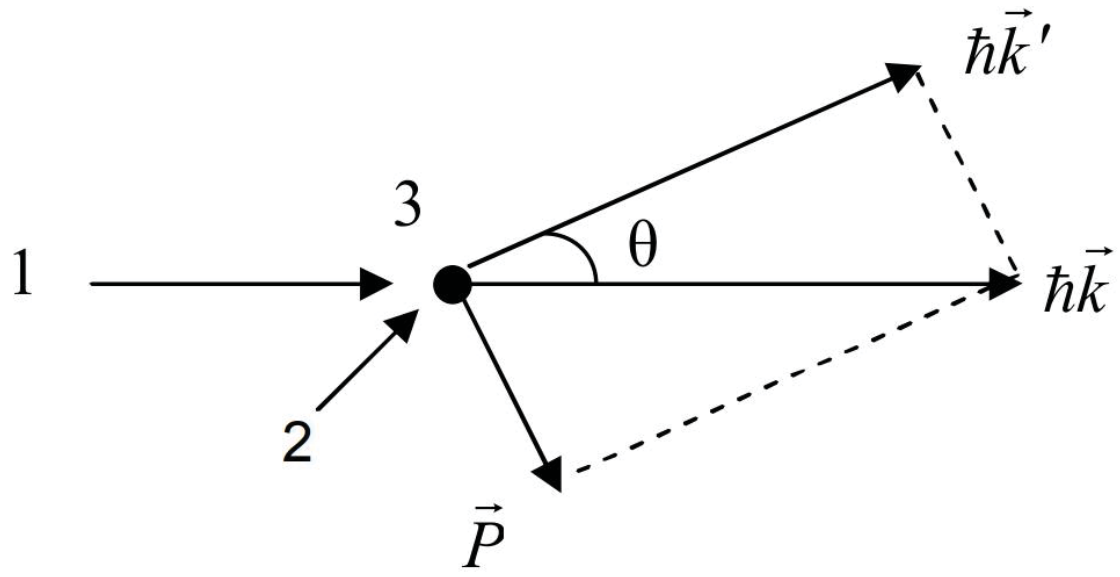
The listed problems and difficulties of quantum field theory make scientists look for other models different from the current ones. One of the most fruitful ideas of recent times is the attempt to combine string and loop theory into a single theory of quantum gravity.

7.7. Typical problems in quantum theory

Problem 1. Based on the Planck hypothesis, prove the formula describing the Compton effect (see Section 7.2.2) of X-ray scattering. Assume that X-rays are a stream of photons with energy and momentum, respectively

$$\mathcal{E} = \hbar\omega,\ P = \hbar k.$$

Solution. Let's assume that the X-ray photon 1 experiences a collision with the electron 2 of the outer shell of the scattering atom 3, which is known to be weakly bound to its nucleus and is quasi-free (see figure).



This means that the energy required for an electron to leave the atom (ionization energy) is many times less than the energy of a photon.

According to the laws of conservation of energy and momentum in this case (see figure)

$$\left.\begin{aligned} \hbar\omega + m_0c^2 &= \hbar\omega' + c\sqrt{P^2 + m_0^2c^2} \\ \hbar\vec{k} &= \vec{P} + \hbar\vec{k}' \end{aligned}\right\}$$

where $\hbar\omega$, $\hbar\omega'$ are the energy of the X-ray photon before and after the collision;

$m_0 c^2$ is the energy of the electron before the collision (at rest);

$c\sqrt{P^2 + m_0^2c^2}$ is the energy of the relativistic electron after the collision;

$\vec{k}$ and $\vec{k}'$ are, respectively, the vectors of the photon wave number before and after the collision;

$\vec{P}$ is the momentum of the electron after the collision;

$\vec{P}_0 = 0$ is the momentum of the electron before the collision;

m_0 is the mass of the electron, $m_0 = 9.1 \cdot 10^{-31}$ kg.

It follows from the first equality that

$$\frac{\hbar(\omega - \omega')}{c} + m_0 c = \sqrt{P^2 + m_0^2 c^2}$$

or

$$\hbar(k - k') + m_0 c = \sqrt{P^2 + m_0^2 c^2}$$

Let's square this expression

$$[\hbar(k - k')]^2 + m_0^2 c^2 + 2\hbar(k - k') m_0 c = P^2 + m_0^2 c^2$$

from which

$$P^2 = \hbar^2 (k - k')^2 + 2\hbar(k - k') m_0 c.$$

Let's square the second equation

$$P^2 = \hbar^2 (k - k')^2$$

The figure shows that

$$P^2 = \hbar^2 (k^2 + k'^2 - 2kk' \cos\theta)$$

Substituting P^2 into the expression obtained from the first equation gives

$$\hbar^2 (k^2 + k'^2 - 2kk' \cos\theta) = \hbar^2 (k^2 + k'^2 - 2kk') + 2\hbar(k - k') m_0 c$$

$$2kk'^{\hbar^2} (1 - \cos\theta) = 2k(k - k') m_0$$

Since

$$\lambda = \frac{2\pi}{k}$$

then finally

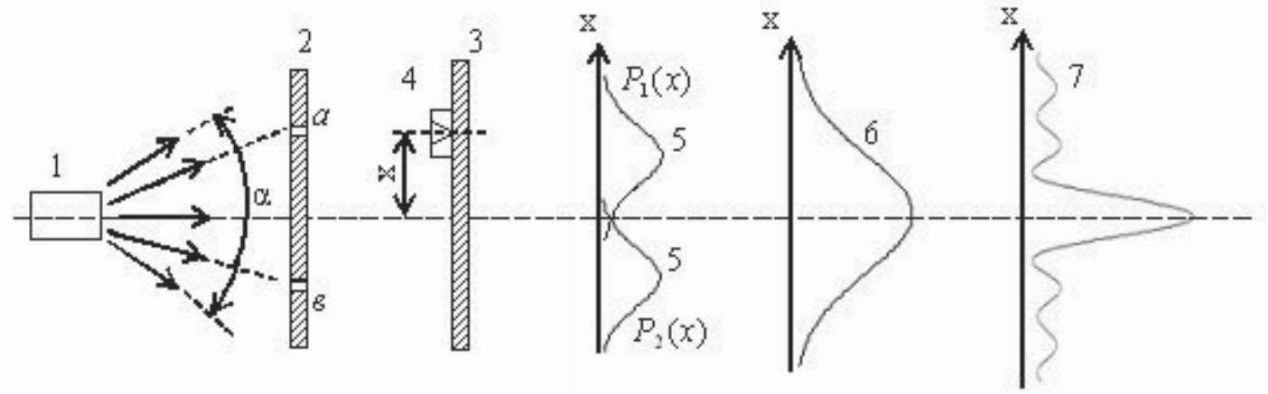
$$\Delta\lambda = \lambda_c (1 - \cos\theta)$$

In the case where the Compton scattering occurs on the electrons of the inner shells (with high binding energy), the photon is not scattered on the electron, but on the atom as a whole with mass M. Then the obtained equation does not change, but

$$\lambda_c = \frac{2\pi\hbar}{Mc}$$

and the value of $\Delta\lambda$ (Compton displacement) is M / m_0 times smaller.

Problem 2. (Feynman's problem). Electron gun 1 (see figure) emits a beam of electrons with the same energy, scattered at an angle α.



The beam passes through a thin metal plate 2 with two small holes a and b. Behind it there is another plate 3, which absorbs electrons. In front of the plate 3 at a distance x from its axis a detector 4 is suspended, such as a Geiger counter, which counts the number of electrons caught in it. By moving the detector along the plate 3 and counting the number of electrons P(x) per unit time for different values of x, the curves are plotted:

- P_{12} (x) when both holes a and b are open;

- P_1 (x) when only one hole a is open;

- P_2 (x) when only hole b is open.

Here P(x) is the probability distribution of electrons hitting a given point of the corresponding screen in the x direction, which is

defined as the ratio of the number of electrons hitting a given point to the number of all electrons hitting the screen.

It is required to do following:

1) plot curves $P_1(x)$, $P_2(x)$, and $P_{1,2}(x)$ on graphs combined with the figure;

2) establish dependences $P_{1,2}(x)$, $P_1(x)$, and $P_2(x)$ and explain their nature;

3) determine how the established dependence changes by including in the experiment an observer who can follow the electrons and find out through which hole the electron that is currently registered by counter 4 has passed (for example, by flashes of light near the holes).

Solution.

1. When electrons pass only through one of the holes and the other is closed, they behave like classical particles - corpuscles (for example, bullets fired from a machine gun installed instead of electron gun 1). In this case, the functions $P_1(x)$ and $P_2(x)$ have the Gaussian law distribution of a random variable (scatterplot) shown in the figure (curves 5).

2. If electrons were in fact classical particles, it is quite obvious that for them

$$P_{1,2}(x) = P_1(x) + P_2(x), \qquad (1)$$

i.e., the distribution curve $P_{1,2}(x)$ would have the form of a Gaussian curve (curve 6).

(3) Since, according to quantum theory, electrons obey wave-particle duality, outside of plate 2 their distributions do not simply add up according to (1), but interfere, and, as follows from the Schrödinger equation, their probability of being at a given point x, determining functions $P_1(x)$, $P_2(x)$, and $P_{1,2}(x)$, is proportional to the square of the amplitude of the corresponding wave functions

$\varphi_1(x)$, $\varphi_2(x)$, and $\varphi_{1,2}(x)$.

In other words,

$$P_1(x) = |\varphi_1(x)|^2;\ P_2(x) = |\varphi_2(x)|^2;\ P_{12}(x) = |\varphi_1(x) + \varphi_2(x)|^2. \quad (2)$$

From relation (2) it follows therefore that

$$P_{12}(x) = |\varphi_1(x)|^2 + |\varphi_2(x)|^2 + 2|\varphi_1(x)||\varphi_2(x)|. \quad (3)$$

or

$$P_{12}(x) = P_1 + P_2 + 2\sqrt{P_1 P_2}\cos\delta. \quad (4)$$

where δ is the phase difference for the wave functions, φ_1 and φ_2, which is determined by the path difference and is therefore a function of x. The distribution graph $P_{12}(x)$ is shown in the figure (curve 7) and has the form of an interference curve with alternating maxima and minima. This means that the same electron, passes as a wave, simultaneously through holes a and b, as if splitting into two parts.

4. If, however, an observer is included in the system, he will, of course, still see that the given electron passes either through hole 1 or through hole 2 and no separation occurs. This is because the observation process distorts the true picture of electron behavior because it is accompanied by an electron-photon interaction, which perturbs the behavior of the electron so much that the interference disappears.

Problem 3. Find the solution of the stationary Schrödinger wave equation for a hydrogen-like atom in general form.

Solution. The stationary Schrödinger equation (Section 6.6.3.1) has the form

$$-\frac{\hbar^2}{2m}\Delta\psi + U_p(r)\psi = \mathcal{E}\psi$$

Since in the hydrogen-like atom

$$U_p(r) = -\frac{Ze^2}{4\pi\varepsilon_0 r},$$

then the Schrödinger equation in the spherical coordinate system

$$\frac{1}{r^2}\frac{\partial^2(\psi r)}{\partial r^2} + \frac{1}{r^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\varphi}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi}\right] = -\frac{2m_0}{\hbar^2}\left(\mathcal{E} + \frac{Ze^2}{4\pi\varepsilon_0 r}\right)\psi$$

We are looking for a solution in the form of

$$\psi\,(r, \theta, \varphi) = \psi_1\,(r)\,\psi_2\,(\theta, \varphi),$$

In this case, the original equation splits into two equations for the functions ψ_1 and ψ_2, namely

$$\frac{1}{r}\frac{d^2(r\psi_1)}{dr^2} = -\frac{2m_0}{\hbar^2}\left(\mathcal{E} + \frac{Ze^2}{4\pi\varepsilon_0 r}\right)\psi_1 \cdot \frac{1}{\sin^2\theta}\frac{\partial^2\psi_2}{\partial\varphi^2} + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi_2}{\partial\theta}\right) == -\frac{2m_0 r^2}{\hbar^2}\left(\mathcal{E} + \frac{Ze^2}{4\pi\varepsilon_0 r}\right)\psi_2.$$

The first is an ordinary homogeneous differential equation of the second order with variable coefficients. It allows for a fairly simple solution. The solution of the second equation is sought by analogy with the general solution in the form of

$$\psi_2(\varphi, \theta\,) = \psi'_2(\varphi)\psi''_2(\theta)$$

Let us dwell in more detail on the solution of the first equation. This equation is simplified and written in the form

$$\frac{d^2\xi}{d\rho^2} - \left(\sigma - \frac{2}{\rho}\right)\xi = 0$$

by means of substitution

$$\xi = \rho\psi_1, \rho = \frac{r}{r_1}, \sigma = \frac{\mathcal{E}}{\mathcal{E}_1},$$

where r_1 and $\mathcal{E}_1$ are, respectively, the values of these quantities for the first Bohr orbit, namely

$$r_1 = \frac{4\pi\varepsilon_0\hbar^2}{m_0 e^2};\ \mathcal{E}_1 = \frac{m_0 e^4}{32\pi^2\varepsilon_0^2\hbar^2}.$$

The resulting equation is solved by substitution

$$\xi(\rho) = e^{\alpha\rho} q(\rho),$$

and $q(\rho)$ is found as a power series

$$q(\rho) = \sum_{i=1}^{n} a_i \rho^i,$$

where α is an arbitrary integration constant.

Due to its arbitrariness, it is chosen so that

$$\alpha^2 = -\sigma > 0.$$

The coefficients of a_i in the power series $q(\rho)$ are arbitrary constants. Substituting $\xi(\rho)$ into the original equation gives the following recurrence formula for the coefficients a_i

$$a_{i+1} = \frac{2(\alpha i - 1)}{i(i+1)} a_i$$

In order to get all the coefficients a_i from this formula, one of them, for example a_1, must be given. It is common to set $a_1 = 1$. As the order of a_i increases, i.e., as i increases, when $i >> 1$

$$a_{i+1} \approx \frac{2\alpha}{i} a_i$$

This means that

$$\left.\begin{aligned} a_2 &= \frac{2\alpha}{2}\cdot 1 \\ a_3 &= \frac{2\alpha}{3}\cdot\frac{2\alpha}{2}\cdot 1 \\ a_4 &= \frac{2\alpha}{4}\cdot\frac{2\alpha\cdot 2\alpha}{1\cdot 2\cdot 3} \\ &\dots \\ a_{i+1} &= \frac{(2\alpha)^i}{i!} \end{aligned}\right\}$$

From the theory of series (see Appendix 2) it follows that

$$e^{2\alpha\rho} = \sum_{i=1}^{\infty}\frac{(2\alpha)^i}{i!}\rho^i.$$

Substitution yields

$$e^{2\alpha\rho} = \sum_{i=1}^{\infty} a_i\rho^i;\ \ \psi_{11}(\rho) = \frac{1}{\rho}e^{\alpha\rho}.$$

Let us then choose an arbitrary constant α in such a way as to provide the obvious condition that follows from the possible real behavior of the electron inside the atom. It is logical to assume that if $\rho \to \infty$, then $\psi_1(\rho) \to 0$. This condition will always be satisfied if we set $\alpha = 1/n$, and $n = 1, 2, 3, \ldots$

In this case

$$a_{i+1} = \frac{2\left(\frac{i}{n}-1\right)}{i(i+1)}a_i,$$

$$\psi_{1n}(\rho) \approx \frac{1}{\rho}e^{\frac{1}{n}\rho},$$

and

$$\sigma_n = -\alpha^2 = -\frac{1}{n^2}$$

This means that when $i >> 1$

$$a_i = \frac{2^i}{n^i i!}.$$

In other words, when $i \to \infty$ $a_i \to \infty$. Consequently, $\xi(\rho)$ for large α is finite and

$$\psi_1(\rho) = \frac{\xi(\rho)}{\rho}$$

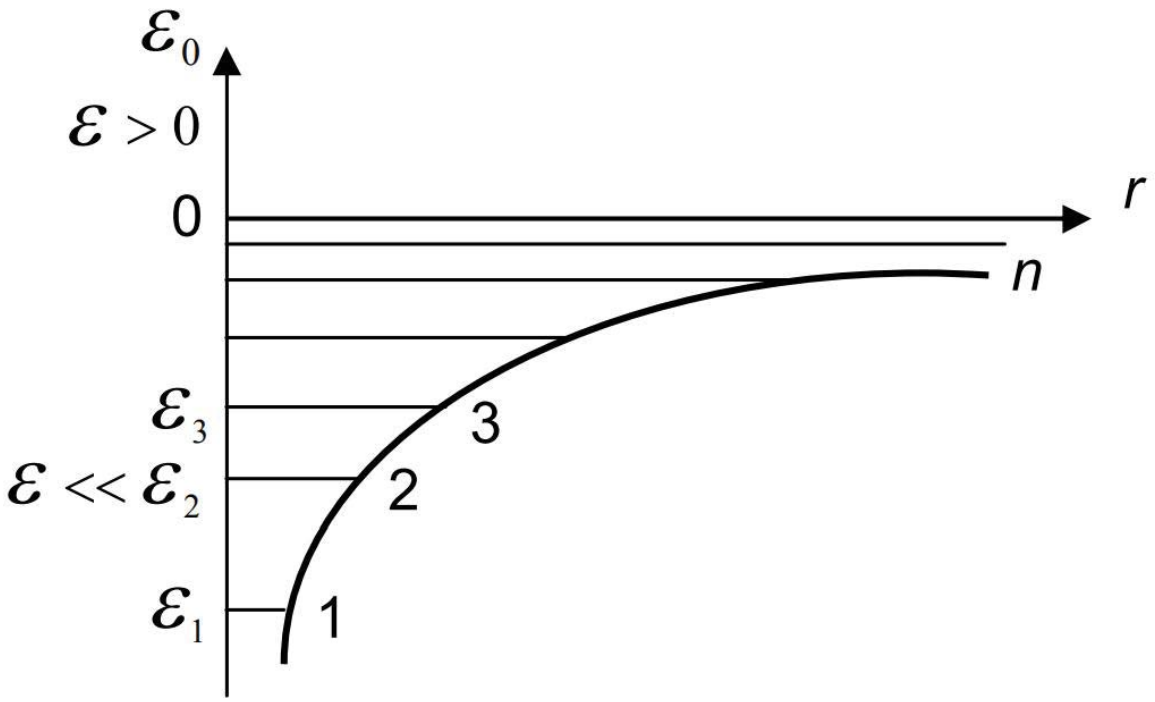
at large α tends to zero.

On the other hand, from the condition of energy quantization

$$\mathcal{E}_n = -\frac{1}{n^2}\mathcal{E}_1$$

The number n sets the allowed discrete values of the energy of the electron bound in the atom. It characterizes the so-called energy levels of this electron. It is also known that the principal quantum number determines the state of the electron inside the atom and is consistent with experiment as well as Bohr's postulates.

Thus, when $\mathcal{E} < 0$ (the electron is bound inside the atom), the energy spectrum of the electron is a discrete sequence of energy levels, visually depicted in the figure. As can be seen from the figure, the energy spectrum of the electron bound in the atom in the quantum-mechanical view differs significantly from the energy spectrum in the classical view (see Figure 7.10).



In the general case of the hydrogen-like atom, the electron state inside the atom is determined not only by the radial spherically symmetric wave function $\psi_1(r)$, but also by the spherical wave function $\psi_2(\theta, \varphi)$.

The solution of this function is found by analogy with the general solution, namely

$$\psi_2(\theta = \psi'_2(\varphi)\ \psi''_2(\theta),$$

For the functions ψ'_2 and ψ''_2 after substitution we obtain

$$\frac{d^2\psi_2'(\varphi)}{d\varphi^2} = -\frac{2m_0r^2}{\hbar^2}\left(\mathcal{E} + \frac{Ze^2}{4\pi\varepsilon_0 r}\right)\sin^2\theta\psi_2'(\varphi),$$

$$\frac{d}{d\theta}\left(\sin\theta\frac{d\psi_2''(\theta)}{d\theta}\right) - \frac{2m_0r^2}{\hbar^2}\left(\mathcal{E} + \frac{Ze^2}{4\pi\varepsilon_0 r}\right)\psi_2''(\theta) = 0.$$

The solution of these equations (given without derivation) gives

$$\psi'_2(\varphi) = e^{jm_l\varphi},$$

$$\psi_2''(\theta) = c(1-x^2)^{\frac{|m_l|}{2}} \frac{d^{l+|m_l|}}{dx^{l+|m_l|}}(x^2-1)^l,$$

$$x = \cos\theta$$

$$(1-x^2)^{\frac{|m_l|}{2}} = \sin\theta^{|m_l|}$$

$$(x^2-1)^l = -\sin^{2l}(\theta)$$

Problem 4. Calculate, in first approximation, the energy value of the ground state of the hydrogen atom.

Solution. This problem is solved according to the semi-classical theory, which does not use the Schrödinger wave equation, but proceeds from a combination of the laws of classical and quantum mechanics.

This approach makes it possible to find a solution to a number of problems of quantum mechanics. These solutions are approximate, because the combination of the laws of classical and quantum mechanics is approximate and admissible only within certain limits. In addition, uncertainty relations are imprecise equations and allow only an order of magnitude to be estimated. The application of semi-classical theory to calculate, in first approximation, the parameters of the ground state of the hydrogen atom and hydrogen-like atoms is acceptable in principle. Indeed, on average, the hydrogen atom remains in the ground state for quite a long time (almost infinitely long), and from this point of view obeys the laws of classical mechanics with corrections introduced by the uncertainty relations in the ground state.

From the classical point of view, the total energy $\mathcal{E}$ of the hydrogen atom (hydrogen-like atom) in the ground state, i.e. in the conditionally stationary orbit, is defined as the sum of the kinetic energy of the rotating electron of the atom

$$\mathcal{E}_k = \frac{m_e v^2}{2} = \frac{P^2}{2m_e}$$

and the potential energy of its Coulomb interaction with the nucleus

$$U = -\frac{e^2}{4\pi\varepsilon_0 r}$$

where m_e is the mass of the electron;

r is the orbital radius of the electron.

It is, in this way,

$$\mathcal{E} = \frac{P^2}{2m_e} - \frac{e^2}{4\pi\varepsilon_0 r} \quad (1)$$

Let us now introduce a correction to this classical equation derived from the uncertainty relation for momentum

$$\Delta P\, \Delta r \approx \hbar \quad (2)$$

As a first approximation, we can assume that in the general case the uncertainties ΔP and Δr are determined from the assumption that

$$\Delta P \cong P,$$

Substitution in the uncertainty relation (2) gives

$$P = \frac{\hbar}{r} \quad (3)$$

Assuming the approximate equality (3) to be strict, we substitute it into (1) and obtain

$$\mathcal{E} = \frac{\hbar^2}{2m_c r^2} - \frac{e^2}{4\pi\varepsilon_0 r} \quad (4)$$

On the other hand, it is known that in the ground state the system is at the lowest possible energy level (n = 1). This means that in this state there must be a condition of minimum energy

$$\frac{d\mathcal{E}}{dr} = 0 \tag{5}$$

From (4) and (5) it follows therefore that

$$-\frac{\hbar^2}{m_e r_0^3} + \frac{e^2}{4\pi\varepsilon_0 r_0^2} = 0 \tag{6}$$

where r_0 is the radius of the orbit in the ground state. This radius is called the radius of the first Bohr orbit in honor of Niels Bohr, who first calculated it based on the postulates of the theory of the atom.

It follows from (6) that

$$r_0 = \frac{4\pi\varepsilon_0\hbar^2}{m_e e^2}. \tag{7}$$

Substituting the values of constants (Appendix 6) in equations (4) and (7) we obtain for the energy of the ground state of the atom in the first approximation the value

$$\mathcal{E}_0 = \frac{m_e e^4}{32\pi^2\varepsilon_0^2\hbar^2} \cong -13.6\ eV.$$

Problem 5. Show that the results of the calculations performed in Problem 4, based on the uncertainty relation, directly follow also from the postulates of Bohr theory.

Solution. According to Bohr's semi-classical theory, the atom corresponds to Rutherford's model.

From the quantization condition proposed by Bohr in accordance with his postulates (see section 7.2.4 and equations 7.53), it follows that the angular momentum of the electron

$$m_0\, \upsilon\, r_n = n\, \hbar, \quad n = 1, 2, 3, \ldots$$

where r_n is the radius of the nth orbit of the electron.

From here

$$r_n = \frac{n\hbar}{m_o \upsilon}$$

Since, according to Rutherford's model of the atom for a stationary orbit, the Coulomb force of attraction of the electron by the nucleus is balanced by the centripetal force, then

$$\frac{m_0 \upsilon^2}{r_n} = \frac{Ze^2}{4\pi\varepsilon_0 r_n^2},$$

where Z is the charge of the nucleus

From here

$$r_n = \frac{4\pi\varepsilon_0 n^2 \hbar^2}{Z m_o e^2}$$

For a hydrogen atom in the ground state $Z = 1$ and $n = 1$, $m_0 = m_e$, therefore

$$r_n = \frac{4\pi\varepsilon_0 \hbar^2}{m_o e^2}$$

which corresponds to equation (7) of Problem 4.

Problem 6. Find the wave function of a microparticle localized in a three-dimensional rectangular potential well, see Figure 6.29 (potential box) with sides l_1; l_2; l_3.

Solution. A microparticle in a potential well is in a stationary state, which is described by the stationary Schrödinger equation

$$-\frac{\hbar^2}{2m}\Delta\psi(x,y,z) + U_p(x,y,z) = \mathcal{E}\psi(x,y,z), \qquad (1)$$

where $\psi(x, y, z)$ is the desired wave function.

In the stationary state defined by the well,

$$U_p(x, y, z) = \begin{cases} 0 \text{ - for any (x, y , z) inside the well, including on its walls} \\ \infty \text{ - for any (x, y , z) outside the well.} \end{cases} \quad (2)$$

In the Cartesian coordinate system

$$\Delta\psi(x, y, z) = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}.$$

Equation (1) outside the well

$$\Delta\psi - \alpha^2\psi = 0, \quad (3)$$

where

$$\alpha^2 = \frac{2m}{\hbar^2}\left(U_p - \mathcal{E}\right) \to \infty \quad (4)$$

The finiteness condition (3), taking into account (4), is provided only under the condition

$$\psi(x, y, z) = 0, \quad (5)$$

Equation (5) is the solution to the problem outside the well. Inside the well

$$\Delta\psi + k^2\psi = 0, \quad (6)$$

$$k^2 = \frac{2m\mathcal{E}}{\hbar^2}. \quad (7)$$

The boundary conditions of the problem in this case are the requirements following from (2), according to which on the pit walls $\psi(x, y, z) = 0$, i.e.

$$\left.\begin{array}{c} \psi(0, y, z) = \psi(x, 0, z) = \psi(x, y, 0) = 0 \\ \psi(l_1, y, z) = \psi(x, l_2, z) = \psi(x, y, l_3) = 0 \end{array}\right\} \quad (8)$$

We look for a solution of equation (6) in the form

$$\psi(x, y, z) = a\,\varphi(x)\,\varphi(y)\,\varphi(z). \quad (9)$$

Substituting (8) into (6) gives

$$\frac{\varphi''(x)}{\varphi(x)} + \frac{\varphi''(y)}{\varphi(y)} + \frac{\varphi''(z)}{\varphi(z)} = -k^2 \tag{10}$$

Relation (10) is possible only when

$$\left.\begin{aligned} \varphi''(x) + k_1^2\varphi(x) = 0 \\ \varphi''(y) + k_1^2\varphi(y) = 0 \\ \varphi''(z) + k_1^2\varphi(z) = 0 \end{aligned}\right\} \tag{11}$$

where

$$k_1{}^2 + k_2{}^2 + k_3{}^2 + = k^2 \tag{12}$$

From (8), taking into account (11), we obtain that

$$\psi\,(x, y, z) = a\,\sin(k_1 x)\,\sin(k_2 y)\,\sin(k_3 z). \tag{13}$$

From (11), given the boundary conditions (7), it follows that

$$\left.\begin{aligned} k_1 = \pm\,\frac{\pi}{l_1} n_1; n_1 = 1, 2, 3 \ldots \\ k_2 = \pm\,\frac{\pi}{l_2} n_2; n_2 = 1, 2, 3 \ldots \\ k_3 = \pm\,\frac{\pi}{l_3} n_3; n_3 = 1, 2, 3 \ldots \end{aligned}\right\} \tag{14}$$

From (12) and (14), taking into account (7), we obtain that

$$\mathcal{E}_{n_{1_1}, n_{2_2}, n_{3_3}} = \frac{\pi^2\hbar^2}{2m}\left(\frac{n_1^2}{l_1^2} + \frac{n_2^2}{l_2^2} + \frac{n_3^2}{l_3^2}\right). \tag{15}$$

Since the square of the absolute value of the wave function in its physical sense determines the probability of detection of a microparticle at a given point of volume V = $l_1 l_2 l_3$, the probability of detection of a particle at any arbitrary point is equal to the sum of probabilities at all points and equals 1, i.e.

$$\int_V |\psi|^2 dV = 1. \tag{16}$$

Equation (16) is called the normalization condition of the wave function.

Substituting in (16) the solution (13) taking into account (15) gives for a constant value of a the value

$$a = \sqrt{\frac{8}{l_1 l_2 l_3}},$$

wherefore finally

$$\psi_{n_1 n_2 n_3}(x, y, z) = \frac{2\sqrt{2}}{\sqrt{l_1 l_2 l_3^2}} \sin\left(\pm \frac{\pi n_1}{l_1} x\right) \sin\left(\pm \frac{\pi n_2}{l_2} y\right) \sin\left(\pm \frac{\pi n_3}{l_3} z\right)$$

Problem 7. Determine the de Broglie wavelength associated with a free-moving electron with energy 10^2 eV.

Solution. According to de Broglie's hypothesis of particle corpuscular-wave duality, each free-moving microparticle with energy $\mathcal{E}$ and momentum P can be compared with a de Broglie monochromatic wave, for which

$$\left.\begin{aligned} \omega &= \frac{\mathcal{E}}{\hbar} \\ \lambda &= \frac{2\pi\hbar}{P} \end{aligned}\right\} \qquad (1)$$

where ω is the cyclic frequency of the wave;

λ is the wavelength.

From the second equation of the system (1) it follows that

$$P = m_e v; v = \lambda\nu; v = \lambda\frac{\omega}{2\pi}. \qquad (1)$$

Substitution in (1) gives

$$\lambda = \frac{2\pi\hbar 2\pi}{m_e \lambda \omega};$$

$$\lambda^2 = \frac{4\pi^2 \hbar}{m_e \omega}.$$

From the first equation

$$\omega = \frac{\mathcal{E}}{\hbar}$$

so

$$\lambda^2 = \frac{4\pi^2 \hbar^2}{m_e \mathcal{E}},$$

from which

$$\lambda = \frac{2\pi\hbar}{\sqrt{m_e \mathcal{E}}},$$

or

$$\lambda = \frac{2\pi\hbar c}{\mathcal{E}};$$

Finally

$$\lambda = \frac{2\pi \cdot 1.05 \cdot^{-34} \cdot 3 \cdot 10^8}{10^2 \cdot 1.6 \cdot 10^{-19}};$$

$$\lambda = 12{,}36 \cdot 10^{-3} \text{ μm}$$

Problem 8. Determine the radius of a water molecule, assuming that the molecule has a spherical shape. Density of water under normal conditions

$$\rho_W \approx 10^3 \text{ kg/m}^3$$

Solution. The water molecule H_2O consists of 10 protons and 8 neutrons (the mass of the electrons is neglected). Hence the mass of one water molecule, $m_W \approx 18 \cdot 1.67 \cdot 10^{-27} = 3{,}006 \cdot 10^{-26}$ kg , and its volume

$$V_W = \frac{18 \cdot 1.67 \cdot 10^{-27}}{10^3} \approx 3 \cdot 10^{-29}\, m^3.$$

On the other hand, $V = \frac{4}{3}\pi r^3$, therefore, the radius of the water molecule

$$r_W = \sqrt[3]{\frac{3V}{4\pi}}$$

$$r_W = \sqrt[3]{\frac{9 \cdot 10^{-29}}{4 \cdot 3.14}} = \sqrt[3]{7 \cdot 10^{-30}} \approx 1.92 \cdot 10^{-10}\, m$$

Problem 9. Determine the value of the Fermi level energy of the metallic sample.

Solution. Depending on the nature of the interactions of atomic nuclei with the electrons of the outer shells, all solids are divided into metals (conductors of electric current) and dielectrics (non-conductors). In metals, the electrons of the outer shell of atoms are weakly bound to their nuclei. In the first approximation, they can be considered free. Since the vast majority of solids have a crystalline structure, the nuclei of their atoms ideally form an ordered periodic crystal lattice. From the energy point of view, a crystal lattice is a set of potential wells separated by potential barriers. Electrons moving through the crystal lattice move from one potential well to another, seeping through potential barriers in accordance with the tunneling effect. The energy of free electrons in metals is high enough and the height of the potential barriers is so low that the electrons move freely throughout the crystal lattice, forming, ideally, a perfect electron Fermi gas. At absolute zero temperature, all

levels below the so-called Fermi level are completely filled with electrons, and above this level are free of electrons.

The wave functions of free electrons of a Fermi gas obey, as is known, the stationary Schrödinger equation

$$-\frac{\hbar^2}{2m}\Delta\psi = \mathcal{E}\psi.$$

Let the metal sample have a cubic shape with side L for simplicity. We look for the solution of this equation in the form of a running plane wave

$$\psi_k(\vec{r}, t) = ce^{j[(\vec{k},\vec{r})-\omega t]},$$

where $\vec{k}$ is the wave vector with components k_x, k_y, and k_z;

$\vec{r}$ is the position vector;

c is a constant.

We find the constant with from the normalization condition

$$\int_V |\psi|^2 dV = 1$$

where $V = L^3$ is the volume of the sample.

By integrating, we get

$$c = \frac{1}{L\sqrt{L}}.$$

It is natural to assume that the wave function at the sample boundary is periodic with period L, i.e.

$$\left.\begin{aligned}\psi(x+L,y,z)&=\psi(x,y,z)\\ \psi(x,y+L,z)&=\psi(x,y,z)\\ \psi(x,y,z+L)&=\psi(x,y,z)\end{aligned}\right\}$$

This means that this system of equations can be considered as boundary conditions for the function $\psi(x, y, z)$.

Substitution gives for quantum numbers of wave vector components

$$k_x=\frac{2\pi}{L}n_x;\ k_y=\frac{2\pi}{L}n_y;\ k_z=\frac{2\pi}{L}n_z,$$

where $n_x, n_y, n_z = 0, \pm 1, \pm 2, \pm 3, \ldots$

Since in the general case for a free electron

$$\mathcal{E}=\frac{P^2}{2m}=\frac{\hbar^2k^2}{2m},$$

then

$$\mathcal{E}=\frac{\hbar^2}{2m}\left(\frac{2\pi}{L}\right)^2\left(n_x^2+n_y^2+n_z^2\right).$$

Thus, the quantum states of free valence electrons in a conventionally cubic metal sample are defined by 4 numbers - n_x, n_y, n_z, and m_s, where m_s is the spin quantum number equal to $\pm\frac{1}{2}$.

Combinations of these numbers take values

$$\left.\begin{aligned}&\left(1,0,0,\pm\frac{1}{2}\right);\ \left(0,1,0,\pm\frac{1}{2}\right);\ \left(0,0,1,\pm\frac{1}{2}\right)\\ &\left(-1,0,0,\pm\frac{1}{2}\right);\ \left(0,-1,0,\pm\frac{1}{2}\right);\ \left(0,0,-1,\pm\frac{1}{2}\right)\end{aligned}\right\}$$

The ground level of the electron corresponds to the minimum energy for $n_x = n_y = n_z = n = 0$ ($\mathcal{E}_0 = 0$). Two electrons with antiparallel

spins are placed on this level. The next level, for which $n^2_x + n^2_y + n^2_z = n = 1$, can be occupied by 12 electrons. Starting from the 15th, the electrons are placed on the second level ($n = 2$). This already accommodates 24 electrons, etc.

Thus, the placement of electrons occurs at increasingly higher levels until all the electrons are used up. At absolute zero temperature, all lower levels will be filled with electrons and all upper levels will be empty.

Let us introduce the phase space of the wave vector k_x, k_y, k_z, which corresponds, as we know, to $(\vec{P} = \hbar\vec{k})$, the momentum $\vec{P}$ space with the axes P_x, P_y, P_z. In this space, all pairs of quantum states of an electron with parallel and antiparallel spins are represented by points, the distance between which is equal to $k_i = \frac{2\pi}{L}$ so that one point has a volume of

$$V = \left(\frac{2\pi}{L}\right)^3.$$

The surface of equal values of energy in this space is equal to

$$\frac{\hbar^2}{2m}\left(k_x^2 + k_y^2 + k_z^2\right) = const$$

and is shaped like a sphere with radius k.

The result is

$$k = \frac{\sqrt{2m\mathcal{E}}}{\hbar}$$

The volume of this sphere is

$$V_k = \frac{4}{3}\pi k^3$$

Since this volume is contained in one point, the number of quantized states is

$$N(\mathcal{E}) = 2\frac{\frac{4}{3}\pi k^3}{\left(\frac{2\pi}{L}\right)^3}.$$

Factor 2 corresponds to two spin values for each electron.

After substituting the values of k and $V = L^3$, we obtain for the number of levels in the energy interval $d\mathcal{E}$

$$N = 8\pi V m\frac{\sqrt{2m\mathcal{E}}}{3(2\pi\hbar)^3}\mathcal{E}.$$

If we combine a surface of equal energy with a surface whose radius corresponds to the Fermi level, then, denoting the concentration of electrons by n, and since $nV = N$, we find

$$\mathcal{E}_F = \frac{12\hbar^2\sqrt[3]{n^2}}{m\sqrt[3]{(3/4\pi)}}.$$

Problem 10. Find the wave function describing the quantum harmonic oscillator and its energy spectrum, see note to Problem 7 of Chapter 6.

Solution. Many problems of quantum mechanics, as already mentioned, are solved based on the method of combination of laws of classical mechanics and equations of quantum mechanic. Let's apply this method to solve this problem.

As we know, in classical mechanics a one-dimensional harmonic oscillator is a particle making harmonic oscillations under the action of an elastic force (see Section 5.1)

$$F = -kx.$$

The potential energy of such an oscillator is

$$U = \frac{kx^2}{2},$$

and the natural frequency of oscillation

$$\omega = \sqrt{\frac{k}{m}},$$

from which

$$U = \frac{m\omega^2 x^2}{2}$$

Substitution into the time-independent Schrödinger equation gives

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + \frac{m\omega^2 x^2 \psi}{2} = \mathcal{E}\psi.$$

The solution of the problem is thus reduced to the solution of the resulting second-order differential equation. To simplify the equation, let us introduce the notation

$$\rho = x\sqrt{\frac{m\omega}{\hbar}}; \quad \alpha = \frac{2\mathcal{E}}{\hbar\omega},$$

then the differential equation takes the form

$$\frac{d^2\psi}{d\rho^2} + (\alpha - \rho^2)\psi = 0,$$

where

$$\psi = \psi(\rho) = f(\rho)e^{-\frac{\rho}{2}};$$

$$f(\rho) = \sum_{k=0}^{\infty} b_k \rho^k,$$

Substituting the solution of $f(\rho)$ into the original equation gives the recurrence formula for the coefficients bk

$$b_{k+2} = \frac{(2k+1-\alpha)}{(k+1)(k+2)} b_k.$$

The resulting solution, however, does not make physical sense, since it is given by a divergent series. Indeed, at $k \to \infty$ $\rho^k \to \infty$ faster than the coefficients b_k to zero. To avoid this divergence, it is necessary to limit the infinite series, that is, to require that, starting from some $k = n$, the coefficients $bn = 0$, that is, that

$$2n + 1 - \alpha = 0.$$

Substituting the value of α gives for the energy spectrum of the oscillator

$$\mathcal{E}_n = \left(n + \frac{1}{2}\right)\hbar\omega, \qquad 0, 1, 2, \ldots.$$

Problem 11. Determine the dependence of the specific electrical conductivity of intrinsic semiconductors on temperature.

Solution. Specific conductivity γ can be determined from Ohm's law, according to which the dependence of the current density vector $\vec{J}$ on the electric field strength $\vec{E}$ is given by the equation

$$\vec{J} = \gamma \vec{E}$$

On the other hand, since the current in the intrinsic semiconductor is the sum of electron and hole currents with densities $\vec{J}_n$ and $\vec{J}_p$, which are created by the motion of carriers with average velocities υ_{nav} and υ_{pav}, and effective masses m_n* and m_p*, then

$$\vec{J} = n_n(-e)\vec{v}_{nav} + n_p e\vec{v}_{pav}$$

and

$$\vec{J} = e^2\left(\frac{n_n\tau_n}{m_n^*} + \frac{n_p\tau_p}{m_p^*}\right)\vec{E}.$$

Since holes and electrons in the intrinsic semiconductor generate and recombine simultaneously with equal probability, their concentrations are equal and proportional to this probability. Thus, comparing this expression with Ohm's law and considering that $n_n = n_p = n$ and denoting

$$e^2\left(\frac{\tau_n}{m_n^*} + \frac{\tau_p}{m_p^*}\right) = \gamma_0,$$

we get that

$$\gamma = \gamma_0 n.$$

Since generation is the result of the transition of an electron from the valence zone to the conduction zone by jumping over the forbidden zone, according to Boltzmann's formula

$$p \approx e^{-\frac{\Delta\mathcal{E}_0}{kT}}$$

Finally

$$\gamma = \gamma_0 e^{-\frac{\Delta\mathcal{E}_0}{kT}}.$$

Problem 12. It is known that with penetration into the depth of the substance, the size of its structural elements and the time of the processes decrease. At the same time, according to the uncertainty relations, the uncertainties of energy increase. Taking the above into account, calculate the time of the processes of electromagnetic and strong

interactions at different structural levels, as well as the energies corresponding to these levels.

Solution. 1. Electromagnetic processes are known to be concentrated at atomic and molecular levels with maximum possible dimensions of structural elements (macromolecules) of 10^{-9}m. Assuming that the speeds of the processes are finite, i.e., equal to the speed of light, we find that

$$\Delta t = \frac{10^{-9}}{3 \cdot 10^{8}} \approx 3 \cdot 10^{-18}\ s.$$

When dimensions and velocities decrease within one or two orders of magnitude (transition to atomic levels and consideration of slow processes associated with the motion of particles), the time intervals may vary within (10^{-15} to 10^{-20}) s. The structural elements are quite strongly connected within atoms and molecules. Their kinetic energies compared to the interaction energies, which are determined by the uncertainty relations, are equal to

$$\Delta \mathcal{E} \approx \frac{\hbar}{\Delta t}.$$

Substitution yields

$$\Delta \mathcal{E} \approx (1 \div 10\ 5\ )\ \text{eV}.$$

2. Strong interactions take place at the nuclear and subnuclear levels.

At the level of nucleon interactions, they occur at a depth of about 10^{-15} m. Reasoning in the same way as in item 1, we find

$$\Delta t \sim 10^{-23}\ \text{s}$$

$$\Delta \mathcal{E} \sim 100\ \text{MeV}$$

At the levels of quark-gluon interactions acting at depths of (10^{-19} to 10^{-27}) m,

$$\Delta t \sim (10^{-27} - 10^{-34}) \text{ s}$$

$$\Delta \mathcal{E} \sim (10^{3} - 10^{10}) \text{ GeV}$$

At the level of the great unification, operating at a depth of (10^{-30} to 10^{-32}) m.

$$\Delta t \sim (10^{-30} - 10^{-32}) \text{ s}$$

$$\Delta \mathcal{E} \sim (10^{14} - 10^{16}) \text{ GeV}$$

Finally, at the level of the assumed action of supergravity at a depth of 10^{-35} m

$$\Delta t \sim 10^{-43} \text{ s}$$

$$\Delta \mathcal{E} \sim 10^{19} \text{ GeV}.$$

Problem 13. In Section 7.6.8 it is shown that from the classical point of view under the action of gravitational collapse any body with mass above a certain value must inevitably and irreversibly be pulled down to a point (singularity). This requirement is limited, however, by uncertainty relations. It follows that the singularity is always surrounded by a Schwarzschild sphere, the radius of which is called the Planck radius, beyond which both any body and any information disappears. This makes us think that the Planck length and its corresponding time defines the minimal possible cell, the quantum of space-time.

What other considerations, besides those mentioned above, are in favor of the space-time quantization hypothesis?

Solution.

It is known that the most significant disadvantages of quantum field theory include divergences, which arise as a consequence of the

assumption of point sizes and charges of elementary particles. This assumption is essentially borrowed from classical theory, which proceeds from the notion of the continuity of space-time and its structures. Continuity of space-time, in turn, means that an infinitely small value of time corresponds to an infinitely small value of space and vice versa. It also means that space and time commute everywhere. On the other hand, it follows from the uncertainty relations that at small space-time scales the classical notion of space-time homogeneity is violated, which indicates that at the same scales its continuity and commutability must be violated. In other words, the idea of point dimensions of space-time should, from the point of view of quantum theory, lead to the possibility of motion and propagation of interactions with unlimited speed, which contradicts the theory of relativity. Under these conditions, it is necessary to introduce a fundamental length l and its corresponding fundamental time t, based on the relation

$$t = \frac{l}{c}$$

The fundamental length is introduced from dimensional theory considerations or from notions of noncommutativity of spatial and temporal intervals. It is interesting to note that in all cases it is equal to the Planck length, i.e. 10^{-35} m.

From gauge symmetry of quantum field theory it follows that there must also exist a fundamental non-point charge, reflecting, somehow, internal discreteness and orders of elementary particles.

All this led to generalizations of quantum field theory called non-local field theory, i.e. field theory that does not work in a point.

Note. The theory of dimensions proceeds from the fact that if it is known on which quantities the sought quantity depends, then we can make an equation of dimensions, the left part of which contains the dimensionality of the sought quantity and the right part contains the

dimensionality of the constants, i.e. the quantities on which the sought quantity depends, raised to the power of x, y, z,... By further equating the indices of the powers of the corresponding dimensions on the left and right, we form a system of equations from which the unknowns x, y, z,... are found. The dimensionless coefficients are selected if necessary, based on various theoretical and experimental data.

Let us illustrate this by the example of the derivation of the fundamental length formula in quantized space-time, which is the subject of the present problem.

Let us take as a starting point the postulate that the fundamental length l is a function of the universal constants that determine the physical quantities bounding it at the Planck depth, namely:

- gravitational constant G; at the Planck depth, as indicated, the electrically weak interaction compensates the strong one, with the result that gravity is the determinant here;

- Planck's constant, which determines the effect of the uncertainty relation;

- the speed of light, which determines the energy of the particles.

In other words,

$$l = f(G, \hbar, c)$$

or

$$[L] = [G]^x [\hbar]^y [c]^z,$$

where

$$[G] = \mathrm{m}^3\ \mathrm{kg}^{-1}\ \mathrm{s}^{-2}$$

$$[\hbar] = \mathrm{kg}\ \mathrm{m}^2\ \mathrm{s}^{-1}$$

$$[c] = \mathrm{m}\ \mathrm{s}^{-1}$$

Thus, under the conditions of the indicated dimensions

$$L = (L^3 m^{-1}T^{-2})^x (L^2 mT^{-1})^y (LT^{-1})^z.$$

Let's make a system of equations

$$\left.\begin{array}{c} 3x + 2y + z = 1 \\ -x + y = 0 \\ -2x - y - z = 0 \end{array}\right\},$$

from which

$$x = y = \frac{1}{2}$$

$$z = -\frac{3}{2}$$

Thus, after substitution we get

$$l = \sqrt{\frac{G\hbar}{c^3}}$$

The obtained value coincides completely with the formula for the fundamental length derived by another method (see Section 7.6.8).

Appendix to Part 5

Problems for chapter 7

Problem 132. How many photons n is emitted by a helium-neon laser of wavelength λ = 630 nm with power N = 5 mW in time t = 10 s in continuous mode?

Problem 133. How will the wavelength of X-rays change as it interacts with matter? Why do these interactions result in electrons escaping from the substance? At what speed and kinetic energy do the electrons fly out? At what angle in relation to the incident radiation do they fly out?

Problem 134. Nuclear fusion reactions, as we know, can only take place at very high temperatures. For this reason they are also called thermonuclear. Explain why these reactions require high temperatures. Calculate the average temperature at which it is theoretically possible to carry out fusion reactions of the hydrogen-helium cycle. In reality, this temperature may be lower. Explain why.

Problem 135. How do you calculate the energy and specific bonding energy of an atomic nucleus? Show by the example of the aluminum nucleus $^{27}_{13}Al$.

Problem 136. From the point of view of quantum theory, the behavior of the electron is described by the wave function, which is a solution of the Schrödinger equations. However, in some cases it is acceptable, with a good enough approximation, to consider the electron as a semiclassical particle, which, on the one hand, moves inside the atom in a circular orbit around a stationary nucleus, and, on the other hand, can be represented as a De Broglie wave. It is required to calculate the radius r_0 of the electron orbit in the ground state of the hydrogen atom, the

mechanical energy W, and the ionization energy of the atom, taking into account the above approximation.

Problem 137. Determine the electrical conductivity of intrinsic and doped semiconductors.

Problem 138. Before the discovery of neutrons by Sir James Chadwick in 1932, it was assumed that the atomic nucleus consisted of protons and electrons. However, this assumption led to the so-called "nitrogen disaster" because it conflicted with the experiment. Explain the essence of the nitrogen disaster and how it was solved with the discovery of neutrons.

Problem 139. Can a neutrino laser be built?

Problem 140. Monochromatic light polarized in the vertical plane is incident on a polaroid whose optical axis forms an angle φ with the vertical. Will a single photon of this light pass through the polaroid?

Problem 141. It is known that in the helium atom (Z = 2) both electrons are in the ground state at the lowest level, but their spins, according to the Pauli exclusion principle, are antiparallel. And how can we line up their spins in the same direction?

Solving the problems of chapter 7

132. The energy emitted by the laser, $\mathcal{E} = Nt$. The energy of one photon $\mathcal{E}_0 = h\nu = hc/\lambda$. Number of photons emitted by the laser

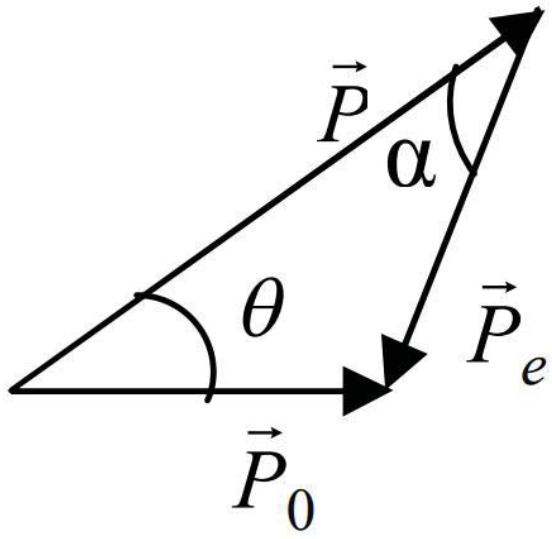
$$n = \frac{\mathcal{E}}{\mathcal{E}_0} = \frac{Nt\lambda}{hc}.$$

Substituting numerical values gives

$$n = \frac{5 \cdot 10 \cdot 630 \cdot 10^{-9}}{6.62 \cdot 10^{-34} \cdot 3 \cdot 10^{8}} = 1.6 \cdot 10^{17}.$$

133. As a result of interaction with matter, radiation deviates from the direction of its propagation by a certain angle θ. Radiation photons transfer part of their energy to the free electrons of matter. As a result, the frequency of radiation decreases and the wavelength λ increases by Δλ. Part of the electrons, which increased their energy due to radiation, increase their kinetic energy and fly out of the substance at an angle α to the direction of the incident radiation. The considered process of interaction of radiation with matter is called Compton scattering, and the angle θ is the Compton scattering angle. The current created by the escaped electrons is called a photoelectric current, and the process of the creation of a photoelectric current by irradiating a substance is called a photoelectric effect.

Compton scattering is described using equations for the laws of conservation of momentum and energy. Consider the interaction of an electron of mass m0 and a photon of radiation with energy and momentum W_0, $\vec{P}_0$ before scattering and W, $\vec{P}$ after scattering, respectively. For simplicity, let us assume that in the initial state (before scattering) the electron was at rest, and its final energy and momentum, respectively, were W_e, $\vec{P}_e$. Then according to the laws of conservation of energy and momentum (see figure).



$$m_0c^2 + \mathrm{W}_0 = \mathrm{W}_e + \mathrm{W}$$

$$\vec{\mathrm{P}}_0 = \vec{\mathrm{P}}_e + \vec{\mathrm{P}}.$$

It is obvious that

$$W_0 = \frac{hc}{\lambda}; \quad W = \frac{hc}{\lambda + \Delta\lambda}; \quad W_e = c\sqrt{m_0^2c^2 + P_e^2}$$

(relativistic equation for energy);

$$P_0 = \frac{h}{\lambda}; \quad P = \frac{h}{\lambda + \Delta\lambda}$$

It follows from the figure that

$$P_e^2 = P_0^2 + P^2 - 2P_0P\ \cos\theta.$$

Substitution and the corresponding transformations give

$$\Delta\lambda = \frac{2h^2}{\lambda m_0^2c^2}\sin^2\frac{\theta}{2}.$$

The kinetic energy of the electron is equal to the energy lost by the photon

$$W_k = \frac{hc}{\lambda} - \frac{hc}{\lambda + \Delta\lambda} = \frac{hc}{\lambda + \left(1 + \frac{\lambda^2 m_0^2 c^2}{2h^2 \sin^2\theta}\right)}$$

From the triangle (see figure), according to the Law of sines, it also follows that

$$\sin\alpha = \frac{hc^2\sin\theta}{\upsilon(\lambda + \Delta\lambda)(W_k + m_0c^2)},$$

where

$$v = c\sqrt{1 - \frac{m_0 c^2}{\mathcal{E}^2}}.$$

134. Under certain conditions, the light atomic nuclei combine to form heavier nuclei. This is, for example, the process of fusion of hydrogen isotope nuclei to form helium nuclei. Since the mass of free particles is always greater than the mass of particles bound in the atomic nucleus, a so-called mass defect occurs when the particles are combined, which appears in the form of released intranuclear energy. Once inside the nucleus, protons are mutually attracted and form a stable bond with neutrons, even though they simultaneously experience strong electrostatic repulsion as like-charged particles. This is explained by the fact that, starting from distances comparable with the size of the nucleus, that is, equal to or smaller than Rn = 10^{-14} – 10^{-15} m, the nuclear particles are affected by the intranuclear force of attraction of strong interaction, which is about 100 times greater than the Coulomb force of electrostatic repulsion. Being outside the nucleus, where the nuclear forces are practically zero, the proton is only subject to electrostatic repulsion. In order to get closer to the nucleus, and moreover to penetrate it, the proton must have kinetic energy comparable to the value of its repulsive electric potential U, which creates a potential barrier, which according to the laws of the electrostatic field is described by the equation

$$\vec{F}_E = - \, grad \, U.$$

Hence, with Rn = 10^{-14} m

$$U = \frac{q^2}{4\pi\varepsilon_0 R_n} \approx 0.1 \, MeV.$$

This value is in good agreement with the value obtained experimentally. In order to overcome this potential barrier, the proton must have kinetic energy

$$W_k = \frac{m_p v^2}{2} = U.$$

This means that the proton must be accelerated to a velocity of approximately $4{,}35 \cdot 10^6$ m/s. This velocity corresponds to the average temperature following from the relation

$$W_{kav} = kT_{av},$$

where k is Boltzmann constant ($k = 1.38 \cdot 10^{-23}$ J/K).

Substitution yields

$$T_{av} = 1{,}15 \cdot 10^9 \text{ K}.$$

In reality, the fusion reaction in environments with a large number of particles takes place at lower temperatures. For example, on the Sun, the reaction takes place at an average solar temperature in its interior of $0.4 \cdot 10^7$ K. This is explained by the fact that, according to the laws of distribution, some of the particles have a temperature higher than the average. In addition, according to quantum effects, particles with lower energy leak through a higher potential barrier (tunnel effect).

135. The bonding energy $\mathcal{E}_b$ of an atomic nucleus is determined through the value of the mass defect of the nucleus Δm by the Einstein equivalence formula for mass and energy

$$\mathcal{E}_b = \Delta m c^2$$

The mass defect of a nucleus with atomic number Z, mass number A with neutron mass m_n and their number $N = A - Z$, proton mass m_p and their number Z, electron mass m_e, nucleus mass m_v equal to the mass of a neutral atom m_{at} minus the mass of all its electrons Zm e is determined by the formula (item 7.4.10.2)

$$\Delta m = Z (m_p + m_e) + N m_n - m_{at}.$$

For the aluminum atom

$$Z = 13,\ A = 27,\ N = 14.$$

Substitution yields

$$m_p + m_e = 1{,}00728 + 0{,}0005486 \text{ amu} = 1,\ 0078286 \text{ amu},$$

$$m_n = 1{,}00866 \text{ amu } m_{at} = 26{,}981 \text{ amu}.$$

The value of c^2, expressed in amu, will be

$$c^2 = Ɛ / m = 8.987 \cdot 10^{16} \text{ J/kg} = 931.5 \text{ MeV/amu}$$

Here amu is an atomic mass unit used to express the mass of elementary particles, atoms, and molecules. By definition, 1 amu = 1/12 of the mass of the carbon nuclide ^{12}C. The final result is

$$\Delta m = 13 \cdot 1{,}0078286 + 14 \cdot 1{,}00866 – 26{,}981 = 13{,}1018 – 26{,}9810 = 0{,}242 \text{ amu}$$

$$\mathcal{E}_b = 0{,}242 \text{ amu.} \cdot 931{,}5 \text{ MeV/amu} = 225{,}4 \text{ MeV}.$$

The specific bonding energy is the value $\mathcal{E}_b/A$, i.e., the amount of bonding energy per nucleon. From here

$$\mathcal{E}_b/A = 225.4/27 = 8.348 \text{ MeV/nucleon}.$$

136. The de Broglie wavelength λ as is known (see Section 6.6.3),

$$\lambda = \frac{2\pi\hbar}{m_0 \upsilon}.$$

On the other hand, according to the uncertainty relation for momentum

$$\Delta P \Delta r \geq \frac{\hbar}{2}.$$

Let us assume, without violating generality, that the uncertainty of the position of an electron in an atom is within the radius of its orbit,

and the uncertainty of momentum is within momentum itself. Then the given uncertainty relation by order of magnitude will be

$$P\, r \sim \hbar.$$

It can be assumed that in the ground state, which is characterized by minimal energy and classical behavior, this approximation becomes strict and

$$P = m_0 \upsilon = \frac{\hbar}{r}.$$

Substituting the obtained value of the momentum into the formula for the de Broglie wavelength, we obtain

$$\lambda = 2\pi r$$

In other words, the de Broglie wavelength of the electron in the ground state coincides with the length of the circular orbit of the electron in this state.

The total mechanical energy of an electron in orbit is the sum of the negative Coulomb energy and the kinetic energy of the electron. Since

$$W_k = \frac{m_0 \upsilon^2}{2} = \frac{P^2}{2m_0}; \quad P = \frac{\hbar}{r},$$

then

$$W = \frac{\hbar^2}{2m_0 r^2} - \frac{e^2}{4\pi\varepsilon_0 r}.$$

In the ground state, where the energy is minimal, the condition of its minimum $\frac{dW}{dr} = 0$ applies , from which it follows that in this state

$$r = r_0 = \frac{4\pi\varepsilon_0\hbar^2}{m_0 e^2} = 0.53 \cdot 10^{-10} m; \quad W = W_0 = -\frac{m_0 e^4}{32\pi^2\varepsilon_0^2\hbar^2} = -13.6\ eV.$$

The resulting value of the mechanical energy of the electron expresses the magnitude of the electron's bonding energy in the ground state, as indicated by the minus sign. To ionize an atom, it takes a positive energy equal in magnitude to the bonding energy.

137. The electric current density J of a semiconductor is the sum of the current densities of electronic Je and hole Jp conduction. Expressed in terms of the concentrations of carriers in the conduction and valence zones of the semiconductor's own n_n and n_p, their average velocities υ_n and υ_p, and the carriers' charge *e*, the current density is

$$J = J_n + J_p = n_n (-e)\, \upsilon_n + n_p\, e\, \upsilon_p.$$

The velocities are determined, based on Newton's second law, through the free path times t_n, t_p and the effective masses of carriers m_n, m_p, as well as taking into account the acting Coulomb forces, expressed in terms of the strength of the external electric field $\vec{E}$. The result is

$$\vec{J} = e^2 \left(\frac{n_n t_n}{m_n} + \frac{n_p t_p}{m_p} \right) \vec{E}$$

Since according to Ohm's law

$$\vec{J} = \gamma \overrightarrow{E},$$

where γ is the specific conductivity, then

$$\gamma = e^2 \left(\frac{n_n t_n}{m_n} + \frac{n_p t_p}{m_p} \right).$$

Since in intrinsic semiconductors carriers are born in pairs, their concentrations are equal. On the other side,

$$\frac{t}{m} = \frac{\upsilon}{eE}$$

If we denote $\upsilon/E = k$, then

$$\gamma = en\,(k_n + k_p)$$

The value of k is called the mobility of carriers. For a silicon crystal, for example, $k_n = 0.13$ m^2/V s , and $k_p = 0.047$ m^2/V s.

Carrier concentrations are determined by their generation probability or equal recombination probability. These probabilities, in turn, on the one hand, are proportional to the product of concentrations (electron and hole meeting) and, on the other hand, according to the laws of distribution of quantum statistics (see Section 7.5) are proportional to the exponent of the ratio of the forbidden zone width ΔW_g of a semiconductor to the thermal energy kT of its carriers. In other words,

$$n_n^2 = n_p^2 \sim e^{-\frac{\Delta W_g}{kT}}$$

It follows that

$$\gamma = \gamma_0 e^{-\frac{\Delta W_g}{kT}}$$

$$\gamma_0 \approx e^2 \left(\frac{t_n}{m_n} + \frac{t_p}{m_p} \right).$$

Since the carrier concentration is not the same for doped semiconductors, following can be written for them

$$\gamma = \gamma_{0d} e^{-\frac{\Delta W_d}{2kT}} + \gamma_{0g} e^{-\frac{\Delta W_g}{2kT}}$$

where the d sign refers to the donor levels and the g sign refers to the band gap.

138. The atomic mass of nitrogen is known to be 14. Taking into account that the mass of the electron is almost 2000 times less than the mass of the proton, we conclude that if the nitrogen nucleus consists of protons and electrons, it must contain 14 protons. Since the charge of the nitrogen nucleus Z = +7, it would have to include 7 electrons neutralizing the charge of the nucleus. The remaining 7 electrons, which ensure the neutrality of the nitrogen atom, must have been part of the electron shells of the atom in this case. In total, therefore, a nitrogen atom would have 28 fermions, each with a spin of 1/2. The total spin of the atom would be 28/2 =14, that is, it would have to be integer, and the atom would have to be a boson. From the experiment, however, it followed that the nitrogen atom is a typical fermion. With the discovery of the neutron, this contradiction was automatically resolved. Indeed, it became clear that electrons are not part of the nucleus. Since the mass of the neutron turned out to be approximately equal to the mass of the proton, the nucleus had to consist of 7 protons and 7 neutrons, and the charge-neutralizing 7 electrons had to be part of the electron shells of the atom, which was confirmed in practice. The total spin of the nitrogen atom, consisting of 21 fermions, is fractional and equals 10.5, which means that the nitrogen atom is indeed a fermion. To explain how mutually repulsive protons and neutral neutrons are held in the nucleus, we had to assume that inside the nucleus, in addition to the forces of the electrostatic field, the forces of the intranuclear non-electric field, which was therefore called "strong," act significantly superior to them. It has now been established that the strong interaction does exist and is completely concentrated inside the atomic nucleus. The interaction constant, which characterizes its intensity, for strong interactions was indeed about 100 times greater than for electromagnetic interactions.

139. The laser beam is strictly monochromatic and coherent. In other words, it is formed by a large number of photons of a certain wavelength moving in the same direction. All photons in a laser beam are thus described by the same wave function and are therefore in the same

state. Photons have an integer spin equal to one, that is, they are bosons, to which the Pauli exclusion principle does not apply. That is why no restrictions are placed on the possibility of laser emission by photons. In contrast, the spin of neutrinos is 1/2, so neutrinos are typical fermions and obey the Pauli exclusion principle. This means that in a laser beam of neutrinos there cannot be more than two identical particles with opposite spins. It follows that a beam consisting of a large number of identical neutrinos cannot in principle be obtained, and a beam consisting of different neutrinos will not be coherent and cannot propagate in a strictly specified direction, that is, it will not be laser.

140. It is known that the ratio of the intensity of polarized light passed through the polaroid to the intensity of incident light is proportional to $\cos^2\varphi$. This means that at $\varphi = \pi/2$ the photon will not pass through the polaroid.

141. In order to give the spins of both electrons the same direction, it is necessary to change the state of one of the electrons, for example, to move it from the ground state to the excited state. Otherwise, both electrons would be in the same state, which is impossible because electrons whose spin is half-integer (equal to 1/2) are fermions and therefore, according to the Pauli exclusion principle, cannot be in the same state. To give the spins the same direction, we can place helium atoms in an external magnetic field with induction of $\vec{B}$ and give one of the electrons of their atoms the energy $W = \hbar\omega$. This energy is obtained by the interaction of the magnetic field with the electron's own magnetic moment. Let us recall (see Section 4.2.2.6) that the magnetic moment of a closed current is the value of $\vec{M} = i\vec{S}$, where i is the magnitude of the current, $\vec{S}$ is the vector of the area of the closed loop with current. Proceeding from the semi-classical theory, we will consider the motion of an electron of mass m with charge e inside the atom along a circular orbit of radius r with speed υ. In this case, the electron forms a circular current with force $i = e\upsilon/2\pi r$, whose magnetic moment is equal to

$$\left|\vec{M}\right| = \frac{1}{2} evr = \frac{1}{2m} emvr.$$

The value *m υ r* is nothing but the orbital mechanical momentum of the electron, which in quantum interpretation is quantized and, accordingly, is a multiple of *ħ*. Thus, for the magnetic moment of the electron in the ground state we obtain

$$M = \frac{e\hbar}{2m} = \mu_E = 0.972 \cdot 10^{23}\ Am^2$$

The value μ_E is the so-called Bohr magneton, which, on the one hand, according to Dirac's relativistic theory, is equal to the electron's intrinsic magnetic moment corresponding to spin, and, on the other hand, serves as the unit of magnetic moment in atomic physics and elementary particle physics. When helium atoms are introduced into an external magnetic field, the spins of their electrons are oriented along the direction of the field, so the spin of one of the electrons will flip, and both spins will be oriented in the same direction. Under the influence of the energy of the magnetic field, the electron will go into an excited state, changing its energy by the value ΔW = ħω. If we change the spin direction in a magnetic field with induction B, the energy of the electron will change by ħeB/m or $2\mu_B B$. It follows that the field induction must be greater than or equal to the value of

$$B = \frac{\hbar\omega}{2\mu_E} = \frac{\pi\hbar c}{\lambda\mu_B}.$$

Assuming that the electron emits in the visible spectrum and that, therefore, the wavelength of the emitted light is in the middle of the emission spectrum ($\lambda = 0{,}5 \cdot 10^{-6}$ m), we find that

$$B \geq 2 \cdot 10^4 \text{ Wb/m}^2$$

Israeli Independent Academy for the Advancement of Science

L. Preigerman

M. Brooke

MODERN PHYSICS COURSE

Under the editorship of

Doctor of Philosophy Professor O.E. Baksanskiy

in 5 parts

Third Edition

Appendixes

Recommended by the Academic Council of the Russian State Classical Academy named after Maimonides (Rambam) for undergraduate and graduate students, and for secondary and high school physics teachers.

Moscow URSS 2014

Table of Contents

Appendices

Appendices 1-3 provide the basics of applied mathematics, and Appendix 4 provides problems to all sections of the textbook with solutions.

Appendix 1. Fundamentals of tensor calculus

Tensor calculus is a branch of mathematics that studies the properties of mathematical quantities and includes number theory, algebra, vector algebra, geometry, etc.

1.1. Fundamentals of Number Theory

A **number** is a quantitative measure of a **mathematical quantity**, which, in turn, is defined as a generalized **concept of a measure**. In accordance with this, mathematics distinguishes between **algebraic** quantities or **tensors** and geometric quantities that characterize measures of geometric shapes. A special case of geometric quantities are **trigonometric** quantities.

A distinction is made between **real**, **imaginary** and **complex numbers**. Actual numbers are **integers**, **fractions**, **positive** and **negative**, **rational** and **irrational**. Positive integers arranged in ascending order form a **natural infinite series** (1, 2, 3, ...). Integers can also be **even** and **odd**, **simple** and **composite**. Numbers and the letters expressing them are connected by six **operations** – addition, subtraction, multiplication, division, exponentiation, and extraction of the root. The numbers, the letters expressing them, and the actions connecting them form **algebraic expressions**. The rules for performing actions on numbers and letters are learned in basic school and are given in math reference books. If a mathematical quantity is expressed by one or more numbers, it is called a **tensor**. In other words, a number is a special case of a tensor.

1.2. Geometric quantities

Geometry is a branch of mathematics that studies the shapes and sizes of graphic images, which are called **geometric shapes**. The simplest geometric figure is a point, that is, a figure without shape or size. A geometric figure that has no width or thickness is called a **line**. The line forms a continuous set of points. An infinite number of lines can be drawn through two points. By measuring the distance between these points, moving along each of these lines, it is easy to see that these distances are not equal to each other. The shortest distance between these points is obtained by moving along a single line, which is called a **straight line**. The other lines are **polygonal chain** or **curved line**. A polyline is a line consisting of separate straight sections that do not lie on the same straight line. A straight line bounded on one side is called a **ray**, and a straight line bounded on two sides is called a **line segment**. The segment is measured by **length**. A shape formed by two rays coming from the same point is called an **angle**. The angle is measured in **degrees** or **radians**. A shape that has no thickness is called a **surface**. The surface is formed by a continuous set of lines. A surface that can be aligned with two intersecting straight lines directed at an angle is called a plane. A figure lying in a plane is called a **plane figure**. Flat figures are studied in **planimetrics**. Non-planar, spatial, figures are studied in stereometry, or **solid geometry**. A planar figure formed by a closed polyline is called a **polygon**. The segments of a broken line are called the sides of a polygon. A polygon that lies on one side of each of its sides is called **convex**. The number of sides of a polygon is equal to the number of its interior angles, and the sum of angles of a convex polygon is 180 (n - 2)º. The sum of the angles of a triangle is, respectively, 180º, a quadrilateral is 360º, etc. Lines that lie in the same plane and do not intersect anywhere are called parallel. Quadrilaterals with pairs of parallel sides are called parallelograms, and those with one pair of parallel sides are called trapezoids. A parallelogram with equal sides is called a rhombus. A polygon in which all sides and angles are equal to each other is called a regular polygon. The angles of a regular triangle are, therefore, 60º, a regular quadrilateral is 90º, a regular hexagon is 120º, etc. A 90º angle is

called a right angle, triangles with right angles are called right triangles, and quadrangles are called rectangles. A regular rectangle is called a square. The vertices of polygons (the intersection points of their sides) and the corresponding angles are denoted by the letters A, B, C, D,..., and the sides lying against these vertices by the letters a, b, c, d, ...The sides of a right triangle that form the right angle denoted by C are called the legs (or catheti, singular: cathetus) and are denoted by the letters a and b, respectively. The third side lies against vertex C, is called the hypotenuse, and is denoted by the letter c. The following relations apply to a right triangle

$$a^2 + b^2 = c^2,\ a/c = \sin A,\ b/c = \cos A,\ a/b = \tan A.$$

The first relation is called the Pythagorean theorem. The magnitudes $\sin A$, $\cos A$, and $\tan A$ are called trigonometric. From Pythagoras' theorem it follows that

$$\sin^2 A + \cos^2 A = 1$$

A number of expressions connecting trigonometric quantities follow from the above relations, which can be found in mathematics reference books.

The following applies to non-rectangular (obtuse and acute) triangles

$$a^2 + b^2 \pm 2ab\cos C = c^2$$

Certain relationships are established between numerical and geometric quantities. Each real number corresponds to a point on the number axis, which is a straight line and is formed by a set of points. Since these numbers form a one-to-one set with all the points of the numerical axis that correspond to them, it is called the **real axis**. The antipodes of real numbers are **imaginary** and **complex numbers**. The notion of imaginary numbers arises in the extraction of the square root of a negative real number. An imaginary unit is a number

$$j = \sqrt{-1}. \qquad (1.1)$$

Any imaginary number is represented by a point, which is obtained by rotating the corresponding point of the positive real axis 90º counterclockwise. In other words, the entire set of imaginary numbers fills a straight line perpendicular to the real number axis. This line is called the **imaginary number axis**, respectively. An imaginary number is represented by a real number multiplied by an imaginary unit, respectively, and is represented by a point on the imaginary axis. A complex number $\tilde{z}$ is, respectively, a number represented by the sum of a real and imaginary number as follows

$$\tilde{z} = x + jy. \tag{1.2}$$

The numbers x and y are called the **real** and **imaginary** parts of a complex number or its components. The algebraic sum of complex numbers is found by the rules of addition of real numbers. The real and imaginary parts are added separately to form the real and imaginary parts of the resultant complex number. Multiplication and division of complex numbers are performed according to the rules of operations on real numbers. In doing so, it is assumed that

$$j^1 = j, j^2 = -1, j^3 = j \cdot j^2 = -j, \; j^4 = j^2 \cdot j^2 = +1, j^5 = \ldots$$

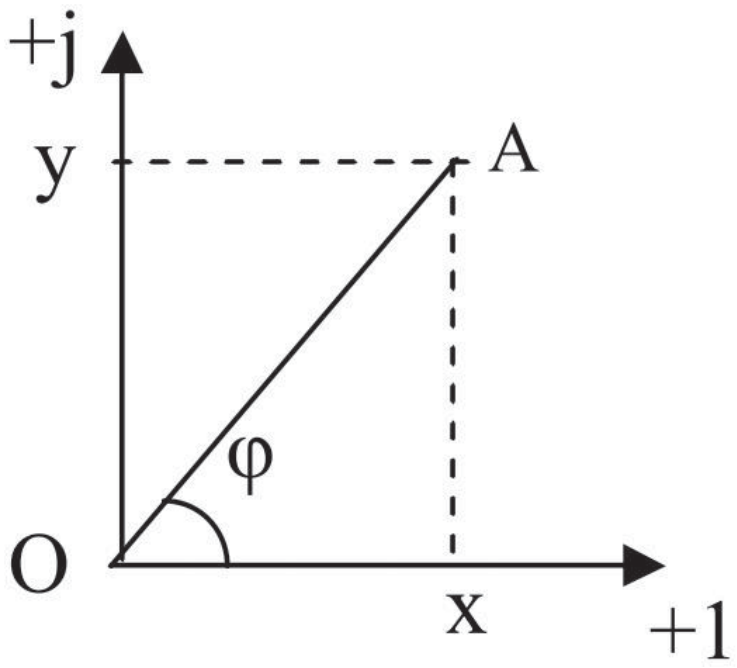


Figure 1.1.

Geometrically, a complex number in two-dimensional space is represented by a point A on the complex plane formed by the horizontal real and vertical imaginary axes (see Fig. 1.1). The quantity

$$z = \sqrt{x^2 + y^2} \tag{1.3}$$

is called the **absolute value** of a complex number. As can be seen from Fig. 1.1, the absolute value of the complex number $z = OA$, where OA is the line segment connecting point A with the origin of coordinates, inclined to the real axis at an angle φ. From Figure 1.1 we can also see that

$$\left.\begin{aligned} x &= z\cos\varphi \\ y &= z\sin\varphi \end{aligned}\right\} \tag{1.4}$$

Substituting (1.4) into (1.2) gives

$$\hat{z} = z(\cos\varphi + j\sin\varphi). \tag{1.5}$$

Below, in Appendix 8, it will be shown that

$$\hat{z} = ze^{j\varphi}, \tag{1.6}$$

where e is an irrational number, which is called the Napier's constant, after John Napier, or Euler's number (not to be confused with Euler's constant), after the Swiss mathematician Leonhard Euler, $e \approx 2.73$.

The expression (1.6) is called the **Euler's formula**.

1.3 Elements of Algebra

Algebra is based on mathematical **operations** on algebraic expressions. Mathematical operations include any action that transforms mathematical quantities and expressions. Combinations of mathematical operations and mathematical signs, such as brackets that determine their order, are called **mathematical operators**. A special case of mathematical operators are algebraic **actions**. Algebraic expressions that are joined by signs of equality, greater than or less than are called **equalities** and **inequalities**, respectively. Equalities are divided into **identities**, which are satisfied for all values of their constituent letters, and **equations**, which are satisfied for some defined values of their

constituent unknown quantities. The values of the unknown quantities that turn an equation into an identity are called the **roots** of the equations. Finding the root is called a **solution** to the equation. A distinction is made between equations with one, two, and many unknown quantities. Equations come in first (linear), second (quadratic), third (cubic), and higher degrees. Equations of the second and third degree are also called quadratic and cubic, respectively. The degree of the equation is equal to the greatest sum of the exponents of the degrees in the unknown terms of the equation. The number of roots of an equation with one unknown quantity is equal to its degree. Solving equations with more than one unknown quantity is impossible. To determine all the unknowns of such equations, it is necessary to solve a system whose number of equations equals the number of unknowns. First-degree equations and quadratic equations, as well as their systems, are studied in elementary school and require no explanation. Cubic equations are solved according to certain rules, which are given in mathematical reference books. There are no rules for solving equations of higher degrees in the general case. The rules for solving equations in individual special cases are given in mathematics reference books. In addition to equations linking algebraic expressions, there are also more complex equations. These include the exponential, logarithmic, irrational, and trigonometric and inverse trigonometric equations.

1.4. Elements of Analytic Geometry

Geometric figures can be described both analytically, by means of corresponding equations, and geometrically, by means of their visual representation in the corresponding coordinate system. The branch of mathematics that studies mathematical objects in terms of unambiguous correspondence between their equations and geometric representations is called **analytic geometry**. In analytic geometry all geometric objects (points, lines, surfaces, bodies) are also given by their equations.

A straight line is given by a first-degree equation in the plane and a system of two first-degree equations in space. Curved lines are also

given by first-degree and higher-degree equations. Lines that are given by second-degree equations are called second-order lines. These include the circle, ellipse, parabola, and hyperbola.

The plane in space is given by the equation of the first degree, surfaces are given by equations of higher degrees. Second-order surfaces are spheres, ellipsoids, paraboloids, hyperboloids, prisms, cones, etc.

Analysis of equations describing geometric objects allows to establish their geometric properties and vice versa.

The position of any point in n-dimensional space can be determined using a system of coordinates. Such a system consists of a set of points in space, the position of each of which is determined by a set of n numbers called point coordinates. The coordinate system is formed by mutually intersecting lines and surfaces called coordinate axes and coordinate surfaces, respectively. Coordinate **axes** are directed relative to each other at equal angles and intersect at one point, which is called the **origin of coordinates** and is usually denoted by the letter O. Coordinate axes are denoted by small letters of the Latin alphabet, e.g. x, y, z, etc. The position of a point in space is defined by projecting it on the coordinate axes. These axes are called Ox, Oy, Oz. The distance of the projection of a point on a given axis to the origin is equal to its coordinate on that axis. The most common is a rectangular **Cartesian coordinate system** consisting of mutually perpendicular lines intersecting at the origin. In addition to the Cartesian system, the cylindrical (polar) and spherical systems, as well as systems of curvilinear coordinates, the axes and coordinate surfaces do not intersect at right angles, are also widely used. A coordinate system consisting of a single line with origin at some arbitrary point is called **linear** and is used to represent a point on a line whose position is determined by a single number. A coordinate system formed by two axes, or in the particular case of the Cartesian system, intersecting at right angles, is used to represent a point in the plane, and is called flat, three axes - spatial, and so on. Equations of basic geometric shapes in different coordinate systems are given in mathematical reference books.

1.5. Basics of tensor and vector algebra

A mathematical quantity characterized by more than one numerical value is called a **tensor**. The tensor is usually denoted by a Latin or Greek letter (in lower- or upper-case) with a curved line as a single sine wave period above it. Each value of a tensor quantity is called its **component** or **constituent**. When writing down a tensor quantity, its components may be specified in parentheses following the tensor notation. The number of components N of the tensor is determined by the size m of the space in which it is given, and by the **rank** n of the tensor using the following formula

$$N = m^n. \tag{1.7}$$

For example, the tensor of rank 2 in three-dimensional space has $N = 3^2$ components. Tensors of rank zero have 1 component in any size space and are called **scalars**. Tensors of the first rank have n components in n-dimensional space and are called **vectors**. Vectors are denoted by letters of the Latin alphabet with an arrow or dash above them. Each tensor component is denoted by a lowercase letter of the Latin or Greek alphabet and has an index at the bottom right whose number of elements equals the rank of the tensor. For example, the components of the second-rank tensor have the following designations - τ_{xx}, τ_{xy}, τ_{xz} ...etc., and the components of the first-rank tensor (vector) - v_x, v_u, v_z. If a scalar is designated alphabetically, the index is not used. The index can be expressed in numbers, letters, etc. Another important characteristic of a tensor is its **absolute value** (modulus), the square of which is expressed through the squares of its components by a simple formula

$$a^2 = \sum_{i,j,\ldots,k} \left(a_{i,j,\ldots,k}\right)^2 \tag{1.8}$$

The absolute value of the tensor is denoted by the same letter, but without its accent or additional signs.

A set of mathematical operations with tensors and vectors is called **tensor algebra** or **vector algebra**.

1.5.1. Geometric representation of the tensor

Tensors can be given a visual geometric representation. In this case, the tensor is represented by a segment of a straight line in N-dimensional space connecting two points in that space, called the beginning and the end of the tensor, with an arrow at the end of the segment indicating the direction of the tensor. The length of this segment is called the absolute value of the tensor or its modulus, and its projections on the coordinate axes are called the tensor coordinates. There is a one-to-one correspondence between the coordinates and the tensor components, i.e. each coordinate corresponds to its tensor component and vice versa. In the particular case of a vector $\vec{a}$, its coordinates are plotted in the chosen scale along the axes Ox, Oy, Oz (Fig. 1.2). The x, y, and z coordinates of a vector are sometimes called **abscissa**, **ordinate**, and **applicate**, respectively. The intersection points of perpendiculars reconstructed from the ends of coordinate segments (vector projections) form the **beginning** and the **end** of the vector.

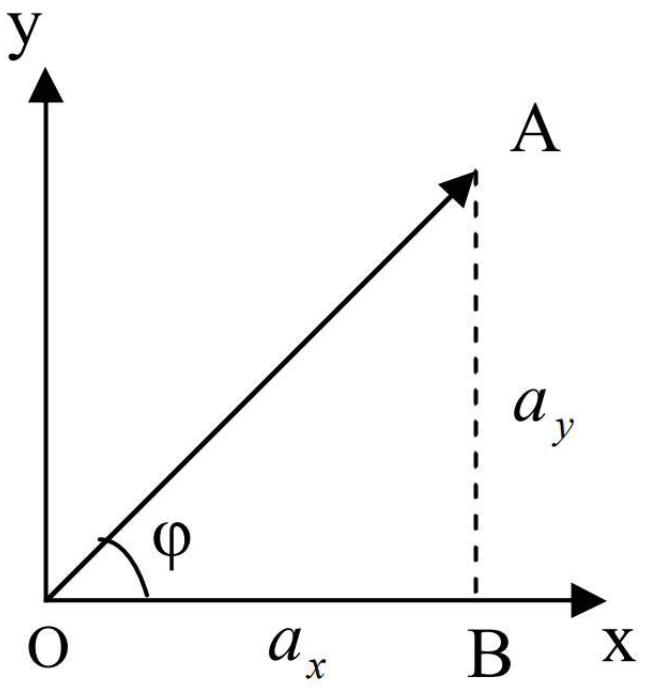


Figure 1.2

Let us consider a right triangle OAB. From this triangle it follows that the modulus of the vector OA is equal to the length of this segment and is found by the formula

$$OA^2 = a^2 = OB^2 + AB^2 = a^2_x + a^2y \quad (1.9)$$

It is easy to see that (1.9) is a special case of relation (1.8) for the tensor. Figure (1.2) also shows that

$$a_x = OA \cos \varphi;\ a_y = OA \sin \varphi. \quad (1.10)$$

From (1.10) and (1.6) it also follows that

$$\hat{a} = ae^{j\varphi} \quad (1.11)$$

As for the tensors of the second and higher ranks, they can be imagined as vectors of the corresponding N-dimensional space in the absence, of course, of the possibility of their visual representation.

1.5.2. Addition of tensors

We will consider the addition of tensors by the example of adding vectors. Let the vectors be given in complex form

$$\hat{a} = a_x + ja_y;\ \hat{b} = b_x + jb_y,$$

then

$$\hat{a} + \hat{b} = (a_x + a_y) + j(b_x + b_y). \quad (1.12)$$

Thus, the addition of two vectors given in the same plane is reduced to the addition of their components. The formulated addition rule is generalized to any number of vectors, including vectors not lying in the same plane, i.e. when adding tensors or vectors given in any dimension space. Let us consider the addition of two or more mutually perpendicular vectors. In a particular case, you can choose a coordinate system in which these vectors are directed along the coordinate axes and are of unit length. In this case, the unit vector along the Ox axis can be written as $\vec{\imath}$ (1,0,0); along the Oy axis as $\vec{j}$ (0,1,0), and along the Oz axis as $\vec{k}$ (0,0,1). Such vectors are called **unit vectors**. Applying the formulated rule of addition

for 3 vectors directed along the coordinate axes, we can write for the resulting vector

$$\vec{a} = a_x\vec{i} + a_y\vec{j} + a_z\vec{k} \qquad (1.13)$$

The resulting ratio is called the rule of **decomposition** of vectors by unit vectors.

Geometrically, the addition of vectors is reduced to the polygon rule, according to which the sum vector is a side that closes a polygon whose sides are equal to the lengths of the added vectors and are parallel to their directions, with the end of each subsequent vector being attached to the beginning of the previous one (see Fig. 1.3). The addition of tensors of any rank is performed according to the same formulated rules.

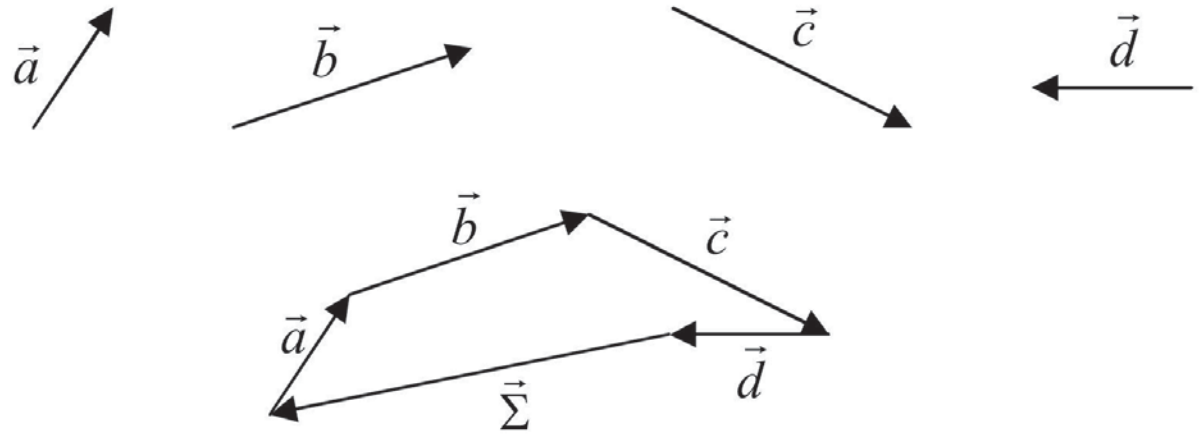


Figure 1.3.

1.5.3. Multiplication of tensors

There is a distinction between the product of a scalar by a vector, the scalar and vector product of two vectors, the mixed and double vector product of three vectors, and the corresponding products of tensors of higher ranks.

The product of a scalar by a vector is a vector whose modulus and coordinates are increased or decreased by some number of times determined by the magnitude of the scalar.

Scalar product of vectors is a scalar equal to the product of the absolute value of one vector and the projection of the other vector onto the first vector. In other words,

$$c = (\vec{a}, \vec{b}) = (\vec{b}, \vec{a}) = ab\cos\varphi \qquad (1.14)$$

where φ is the angle between the vectors.

The scalar product of tensors of higher rank is performed according to the same rule, so that the result is a tensor whose rank is decreased by one unit.

It follows from (1.14) that the scalar product of perpendicular (orthogonal) vectors, e.g. decomposition unit vectors, is zero, and the scalar product of parallel (collinear) vectors is equal to the product of their absolute values. The scalar product of a vector by itself is equal to its square. It follows that, according to (1.13), the scalar product of two vectors when multiplied directly gives the scalar

$$c = a_x b_x + a_y b_y + a_z b_z \qquad (1.15)$$

The vector product of two vectors $\vec{c} = [\vec{a} \times \vec{b}]$ is a vector perpendicular to both vectors being multiplied and equal in absolute value to the area of a parallelogram built on these vectors. This means that the absolute value of the vector product $c = ab\sin\varphi$. The vector product is directed so that, looking from it, you can see the rotation of the first vector to the second at a smaller angle between them counterclockwise (see Fig. 1.4). It follows from this definition that

$$[\vec{a} \times \vec{b}] = -[\vec{b} \times \vec{a}]$$

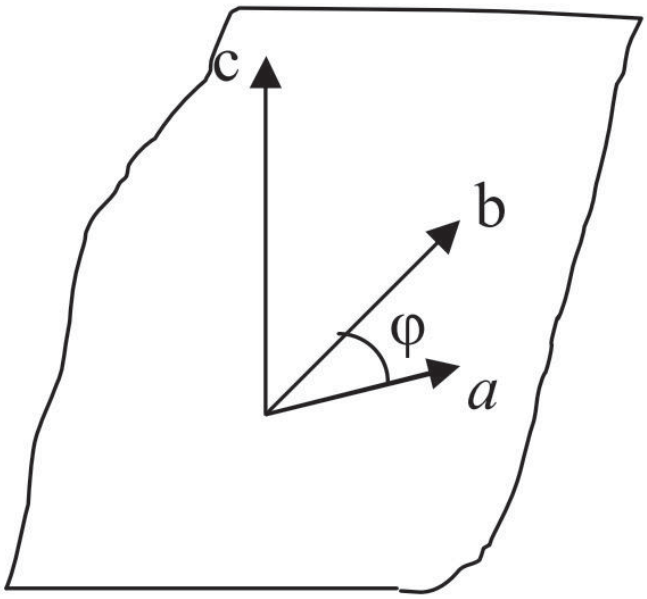


Figure 1.4.

The components of the vector product are found by the determinant rule

$$\left|\vec{a}\times\vec{b}\right| = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}$$
$$= \left(a_x b_z - a_z b_y\right)\vec{i} + \left(a_z b_x - a_x b_z\right)\vec{j} + \left(a_x b_y - a_y b_x\right)\vec{k}$$

The tensor product of two tensors of the same rank is equal to a tensor of the same rank.

The mixed vector-scalar product of three vectors is a scalar equal to the volume of the parallelepiped built on these vectors.

$$\left(\vec{a}\cdot\left[\vec{b}\times\vec{c}\right]\right) = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix} \qquad (1.16)$$

The double vector product of three vectors is found by the formula

$$\left[\vec{a}\times\left[\vec{b}\times\vec{c}\right]\right] = \vec{b}(\vec{a},\vec{c}) - \vec{c}(\vec{a},\vec{b}) \qquad (1.18)$$

There is a mnemonic device to memorize this formula quickly, it reads: "bac minus cab."

Appendix 2. Elements of Mathematical Analysis

All mathematical quantities are divided into **constant** and **variable**. Constant quantities, in turn, are divided into **universal constants**, which retain their values under any conditions, and **conditionally constants**, which retain their values only under certain conditions and within a particular problem. The first values include the irrational numbers known from mathematics like $\pi \approx 3.14$, Napier's constant $e \approx 2.7$. They also include the well-known physical world constants (the speed of light in a vacuum, Boltzmann's constant, Avogadro's number, Planck's constant, the gravitational constant, the mass and charge of the electron, and several others). The conditionally constants are, for example, the speed of uniform rectilinear motion, the current in a circuit with a source of constant electromotive force, the capacity and inductance of conductors, as well as numbers and their alphabetic equivalents, etc.

2.1. Elements of Functional Analysis.

Variables are divided into **independent** variables that change arbitrarily (**arguments**) and **dependent** variables that change when the arguments change. The nature of the change in the dependent variables is determined by the set of mathematical operations performed on the argument. This set is called a **function**. The constants are usually denoted by the first lowercase letters of the Latin or Greek alphabet, as well as by specially assigned letters to these constants, such as π, e, c (speed of light), $\hbar$ (Planck's constant), etc. Variables are usually denoted by the last lowercase letters of the Latin alphabet x, y, z, u, v, w, t, etc. A function is most commonly denoted by the letters y, z, u, etc. or by the signs f, φ, etc.

The statement that y is a function of x is symbolically written as $y = f(x)$ or $y = y(x)$, etc.

The **domain** of a function is the entire set of values it argument can take. The **range** of existence of a function is the set of all values that it takes when the argument changes within its domain. Variables can take a number of discrete values, expressed by certain numbers or, as they say, a countable set. In addition, variables can be continuous and take on any value. If the argument takes discrete values, we speak of a **discrete argument function**. In the other case, we speak of a function of a **continuous argument** or a **function with a connected domain of assignment**. The area of the argument is determined by the conditions of the problem and by the type of function. For example, if a function $y = \sqrt{x-1}$ is given, then the domain of the argument is defined by the expression $x \geq 1$. The range of existence of the function is not limited by anything. On the contrary, for the function $y = x^2$, the domain of the argument is not bounded, and the range of existence of the function is bounded by the relation $y \geq 0$. If a variable in the process of its change tends to zero, it is called **infinitesimal**. A function is called **continuous** at a given point if an infinitesimal change in the argument corresponds to an infinitesimal change in the function. A function that is continuous at all points in the argument domain forms a **connected domain**. The points at which this rule is not observed are called **points of discontinuity**. A distinction is made between functions of **one** variable or **many** variables. A function can tend to infinity, be **bounded** from below, above, or on both sides, and have a **limit**. A function is called **monotonically increasing** if its absolute value increases as the absolute value of the argument increases, and **monotonically decreasing** if its absolute value decreases as the absolute value of the argument increases. The limit of a function y of a continuous argument x at $x \to a$ is its value A, to which the function approaches as close as possible when the argument approaches the value a. This is written down as follows

$$A = \lim_{x \to a} y \tag{2.1}$$

There are different ways to calculate the limit of a function, but the most common is the **L'Hospital's rule**, which we will discuss below. Let us give here, without proof, the three known limits called **special**.

$$\lim_{x\to 0}\frac{\sin x}{x}=1;\ \lim_{x\to\infty}\left(1+\frac{1}{x}\right)^{x}=e;\ \lim_{x\to 1}\sum_{i=1}^{n}\frac{1}{i}-\ln n=C. \qquad (2.2)$$

($e \approx 2.73$, $C \approx 0{,}5772$).

2.2. Elements of Differential Calculus

The limit of the ratio of the increment of a continuous function $\Delta y = y(x + \Delta x) - y(x)$ to the increment of the argument Δx when $\Delta x \to 0$ is called the first derivative of a function, which is denoted dy/dx ($df(x)/dx$) or y', and in the case of the time derivative $\dot{y}$. The definition of the derivative is written as follows

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\frac{dy}{dx}. \qquad (2.3)$$

This differentiation rule can be used to approximate the increment of a function

$$\Delta y \approx y' \cdot \Delta x \qquad (2.4)$$

The first derivative of the first derivative is called the **second derivative**, and the first derivative of the second derivative is called the **third derivative**, etc. The nth derivative is written in the form $d^n y / dx^n$; ($df^n(x)/dx^n$) or as y^{II} (second derivative), y^{III} (third derivative), y^{IV} (fourth derivative), etc. [$f^I(x), f^{II}(x), \ldots f^n(x)$]. The derivatives of a function of one variable are called **total derivatives**. The calculation of derivatives is called **differentiation**. Differentiation is considered the seventh mathematical operation. This action is straightforward and can be performed on any continuous function everywhere. The operation of differentiation is given by the following properties of differentiation:

- the derivative of a constant is zero;

- the derivative of the algebraic sum of functions is equal to the algebraic sum of the derivatives;

- the derivative of the product of a constant by the function is equal to the product of a constant by the derivative of the function;

- the derivative of a product of functions is equal to the sum of the products of the derivatives of each function by all other (non-differentiated) functions;

- the derivative of the function $f\,[\varphi\,(\psi\,(...(x)]$ is equal to the product of the derivative of the external function by the derivative of the first internal function, multiplied by the derivative of the second function, etc.

The rules of differentiation are given in mathematical handbooks and reference books on the basics of mathematical analysis.

L'Hospital's rule for calculating the limit is to write the function $y\,(x)$ in the form of $\frac{y_1}{y_2}$; $y_1 \cdot y_2$; $y_1 - y_2$; $y_1^{y_2}$. This notation reduces to **uncertainties** (that is, expressions that are not computable) of the type $\frac{0}{0}$; $\frac{\infty}{\infty}$; $(\infty - \infty)$; 0^0; ∞^0; 1^∞. Then the L'Hospital's rule is applied, according to which for uncertainties of the first two kinds

$$\lim y = \lim_{x \to a} \frac{\frac{dy_1}{dx}}{\frac{dy_2}{dx}}.$$

For uncertainties of the other types, the L'Hospital's rule is applied after reducing them to the form of the first two uncertainties.

L'Hospital's rule can be applied repeatedly as long as uncertainty persists.

The differentiation of functions of **many** variables is performed, as a rule, on one of the variables under the conditional assumption that the other variables are constant. Such derivatives are called **partial derivatives**. If, for example, a function of two variables $u\,(x, y)$ is given, it has two partial derivatives on x and on y. The partial derivatives are

written as follows: $\frac{\partial u}{\partial x}$; $\frac{\partial u}{\partial y}$. Partial derivatives of higher orders are written in the same way as the first partial derivatives, taking into account the rules for writing total derivatives of higher orders.

The **differential** of a function of one variable is an expression of the form $dy = y' \cdot dx$. The **total differential** of a function of many variables is an expression of the form

$$du = \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy + \cdots \quad (2.5)$$

Each term of this expression is called a **partial differential** or is an approximate partial increment of a function with the increment of the corresponding argument.

2.3. Function assignment. Functional series

A function can be defined in three ways - **analytically**, **in a table**, and **graphically**. The analytical method, in which the function is defined directly, that is, **explicitly** or **implicitly**, by means of a formula containing the entire set of operations that define it, is the most accurate, but the least obvious. The accuracy of the graphical method is much lower, but its clarity is much higher. The tabular method has neither precision nor clarity, but it is often easier to use. The graphical assignment uses the coordinate method discussed in Appendix 1. The argument is plotted on the horizontal axis, and the function is plotted on the vertical axis. Curves of graphs in the plane represent the function of a single variable, and curves in spatial coordinate systems represent the function of many variables. A function of two variables is represented by a curve (diagram) in three-dimensional space. A function of n variables is represented by a curve in (n+1)-dimensional space. The representation of functions as curves allows us to interpret the geometric meaning of the derivative and the differential. In this case, the derivative of a function of one variable at each point is defined as the **slope ratio of the tangent** to the curve at that point. The differential is defined as the increment that the ordinate of the tangent at a given point receives for a given increment of the

argument. Functions of one variable are divided into elementary functions that are expressed in quadrature (that is, using six algebraic operations as well as using exponential, logarithmic, and trigonometric functions without using infinite sums) and special functions that cannot be expressed in quadrature. Elementary functions have derivatives in their entire range of existence, except for the points of discontinuity and the so-called **special** points of curve breaks. Elementary functions include:

- linear functions in which the argument is in the degree not higher than the first;

- integer polynomials of higher powers in which the argument is a positive integer degree; special cases of polynomials are quadratic and fractional-rational functions; the numerator and denominator of these functions are integer polynomials, irrational functions in which the argument is raised to fractional degrees and stands under the radical;

- exponential and logarithmic functions;

- direct and inverse trigonometric functions.

Linear functions are represented graphically by straight lines, polynomials by complex continuous curves with changing character of their increasing or decreasing and changing curvature. The points at which the character of increasing or decreasing function changes are called **extremums** and are subdivided into **maxima** and **minima**. At the maxima, the monotone increase of the function changes to a monotone decrease, and vice versa at the minima. The points at which the curvature changes are called **points of inflection**.

Finding extremum points and inflection points is performed using derivatives, assuming that at extremum points, the first derivatives and at inflection points, the second derivatives are equal to zero.

Functions have, as a rule, besides extremum points and inflection points, also **points of discontinuity**. There are three types of discontinuity: **infinite**, **finite**, and **removable**. At points of the first type,

the function tends to positive or negative infinity. Technically speaking, in this case we are not talking about one, but about an infinite set of points with the same value of the argument. For example, the function $y = 1/x$ tends to acquire an infinite number of infinite values if the value of the argument tends to zero. At finite break points the function jumps from one finite value to another. For example, the function

$$y = \frac{1}{1 + e^{\frac{1}{x-1}}}$$

tends to 1 when the argument x tends to 1 on the left, i.e., on the $x < 1$ side, but when x tends to 1 on the right, the function tends to zero. Finally, in the case of an removable rupture, the function tends to the limit left and right of the discontinuity point equally, but is not equal to its limit at the discontinuity point itself. For example, the limit of a function

$$y = \frac{\sqrt{1+x} - 1}{x},$$

when x tends to zero on the left or right, is 1/2, but at the point $x = 0$ itself, turns into an uncertainty of the 0/0 type. The gap at this point is eliminated by assuming that the specified uncertainty is also equal to 1/2 .

The behavior of the function when the argument tends to infinity is also of great importance. In this case, the limit of the function can be either infinite or finite. For example, it is finite and equal to zero for the function $y = 1/x$ that we considered above.

It is proved that any elementary function can always be expanded into an infinite power series, and any periodically varying function can be expanded into a Fourier trigonometric series. The series into which elementary functions are decomposed are given in reference books on higher mathematics. In general, a power series has the following form

$$f(x) = \sum_{n}^{\infty} a_n (x \pm a)^2, \qquad (2.6)$$

where

$$a_n = \frac{f^{(n)}(a)}{n!};$$

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n.$$

The decomposition of functions into a power series allows us to prove Euler's formula (1.6).

The Fourier series is an infinite series of expansions of a periodic function by the trigonometric functions sin and cos. In the general case a periodic function

$$\varphi(x) = \frac{a_0}{2} + \sum_n a_n \cos n\omega x + b_n \sin n\omega x, \qquad (2.7)$$

where

$$a_n = \frac{2}{T}\int_0^T \varphi(x) \sin n\omega x \; dx\,, \qquad b_n = \frac{2}{T}\int_0^T \varphi(x) \cos n\omega x \, dx$$

An introduction to the expressions by which the Fourier series coefficients are determined is given below.

ω = 2π/T is the angular frequency, and T is the period of change of the initial function φ(x).

2.4. Elements of integral calculus

The operation inverse of differentiation is called **integration**. The problem of integration is to find the function itself, that is, to find a function F(x) whose derivative is equal to the given function f(x). The function F(x) is called a **primitive function**. It follows from this definition that

$$\frac{dF}{dx} = f(x) \text{ or } dF(x) = f(x)\, dx.$$

Since the derivative of an arbitrary constant C is equal to zero, the above equation can also be written in the following form

$$d\,[F\,(x) + \mathrm{C}] = f\,(x)\,dx.$$

A function that is equal to the primitive function with an arbitrary constant is called an **indefinite integral** and is denoted by

$$\int f\,(x)dx$$

In other words,

$$\int f\,(x)dx = F(x) + C \tag{2.8}$$

In mathematics, it is proved that the indefinite integral is expressed by the limit of the infinite sum of infinitesimal quantities with accuracy up to an arbitrary constant. This means that from the geometric point of view integration is reduced to the summation of infinitesimal areas of rectangles with the sides of the ordinate of the function $f\,(x)$ at each point and the element dx of that point. The summation is performed from an arbitrarily chosen value of the argument (coordinate) a to the current coordinate x. The result of the summation is obviously the area of the figure bounded by the curve of the function $f\,(x)$, the Ox axis, and the two ordinates of the curve corresponding to the values of the argument a and x (see Fig. 2.1).

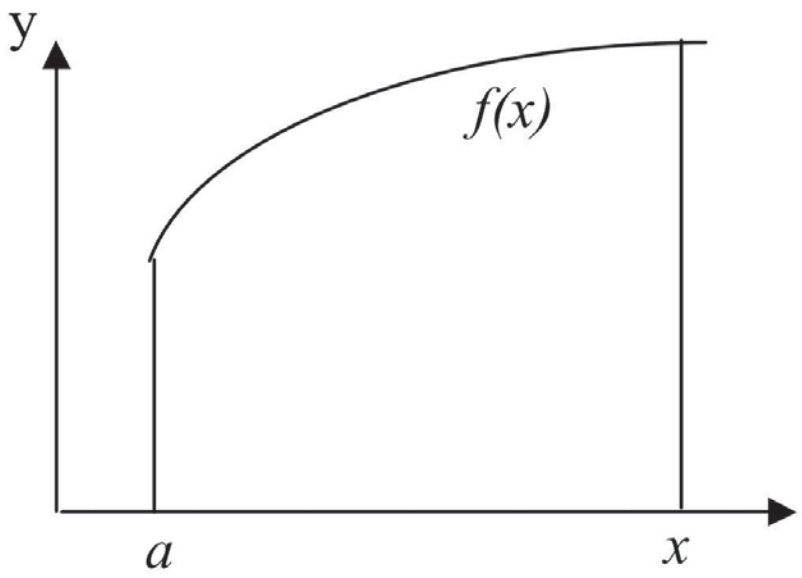


Figure 2.1.

The indefinite integral on the basis of the above can also be determined by means of the so-called limits of integration a and x as follows

$$\int_a^x f\,(x)dx = F(x)$$

Basic properties of the indefinite integral:

- the integral from zero is a constant value;

- the constant multiplier can be taken out of the integral sign;

- the integral of the algebraic sum of functions is equal to the algebraic sum of the integrals of each function;

- from the rule of differentiation of a product of functions follows the so-called rule of integration by parts;

$$\int u(x)dv(x) = u(x)v(x) - \int v(x)du(x)$$

- from the above definition of the indefinite integral follows the basic theorem of the theory of integration, according to which

$$\int_a^x f(x)dx = F(x) - F(a) = F(x) + C \qquad (2.9)$$

The value C is called the **integration constant**.

Integration is often seen as the eighth mathematical operation, which is the inverse of differentiation. Like any inverse operation, integration cannot be performed on any continuous function. Only a relatively small number of functions can be integrated. Integrals from the simplest functions that can be integrated are called **table integrals**. Among the most common methods of integration, in addition to the use of simple table integrals, are methods of integration of rational and fractional-rational functions, integration by parts, as well as the method of substitution (replacement of variables), including trigonometric ones.

Tabular integrals and methods for integrating simple functions are given in the mathematics reference appendices. Some simple rules of integration are also given there. In some cases, approximate methods of integration are also used, e.g., by decomposing a function into a series, etc.

The **definite integral** is an integral with constant limits of integration, namely

$$\int_a^b f(x)dx = F(b) - F(a) \tag{2.10}$$

A geometrically definite integral is interpreted by the area of a flat figure bounded by the curve of the integrand function, the Ox axis, and the ordinates of the function at the points of the limits of integration.

Definite integrals with infinite bounds and integrals from discontinuous functions are called **improper integrals**. The calculation of improper integrals is reduced to the calculation of their limits. To do this, first calculate the indefinite integral, then find the corresponding limits of the obtained functions at infinity or at discontinuity points, and, if these limits exist (i.e., the improper integrals **converge**), then substitute the limits of integration.

In principle, it is also possible to integrate functions of two and many variables.

Consider some integrals from a function of two variables.

2.4.1. Integral depending on a parameter

An **integral depending on a parameter** is a definite integral in which the integrand depends on two variables, and the integration is performed on only one of them. The second variable is called the **parameter**. For example, when integrating over the variable x (y is a parameter)

$$\int_a^b f(x,y)dx = F(y) \tag{2.11}$$

The concept of the integral depending on a parameter is also generalized to integrals of functions of many variables.

2.4.2. Contour integrals

- **integral of the first type**. This is the integral over the length of the curve. It is taken from the segment AB of the curve K, given by the corresponding equation $y = f(x)$. The second variable and the variables of integration are determined from the equation of the curve and from the projections of points A and B of segment AB on the Ox axis and substituted into the required integral, which is thus reduced to a definite integral.

- **integral of the second type**. This is the integral of the projection of the segment AB of curve K on the Ox axis. It is also calculated as a first type integral.

- **an integral of the general form**. This is the integral of the sum of two functions of two variables P(x,y) and Q(x,y) on the parameters x and y and on the projections of the segment AB on the corresponding coordinate axis. The same generalizes to functions of three or more variables. Example of writing

$$J = \int_{AB} P(x,y)dx + Q(x,y)dy \tag{2.12}$$

The contour integral can depend only on points A and B or also on the shape of the path AB. The first case obviously occurs if the integrand is a complete differential of some third function. In the two-dimensional version, for example, if the integrand function

$$P(x, y)\,dx + Q(x, y)\,dy = dU(x, y),$$

then the integral is equal to

$$U(B) - U(A).$$

The full differential condition means, by definition, that

$$P = \frac{\partial U}{\partial x}, Q = \frac{\partial U}{\partial y} \tag{2.13}$$

On the other hand, the existence of a contour integral, as will be shown below, requires that

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} \tag{2.14}$$

This relationship provides a criterion for determining whether the integrand is a complete differential.

The contour integral over a closed path is called a **circulation**. Circulation, which does not depend on the shape of the path, is zero. This fully applies to integrals from functions of three or more variables.

2.4.3. Multiple integrals.

The **two-fold iterated integral** or **double integral** of a function of two variables $f(x, y)$ extended over an area S is the limit of the sum of n elementary areas of rectangles formed by ordinates of the function calculated at points of the area and its elements when they are shrunk to corresponding points, and their number n tends to infinity. It is written as follows

$$\int_S f(x,y)dS = \lim_{\Delta S_i \to 0, n \to \infty} \sum_i^n f(x_i, y_i)\Delta S_i \tag{2.15}$$

The **three-fold iterated integral** or **triple integral**, extended to the volume of the body *V*, is also defined only for a function of three variables given in the volume *V*.

To calculate a multiple integral, for example a double integral, find the equation of the line $y = f(x)$ of the boundary of the area (if it is not given) and determine the limits of the external integral - (a, b) as the projections of the outermost points of the area on the Ox axis, and the

limits of the internal integral as the projections of the outermost points of the boundary line on the Oy axis (f_1 (x), f_2 (x)), and integrate separately on x and y in any convenient sequence. The triple integral is calculated in the same way.

2.4.4. Surface integrals

The surface integral is a generalization of the double integral for a function of three variables given in a connected domain. The surface integral is taken over the area S of a given surface. It is equal to the limit of the sum of the products of the function on the size of the area element when it is pulled down to a point. A distinction is made between surface integrals of the first type, second type, and general type. The differences between them are the same as in the case of contour integrals.

2.5. Elements of differential equations

An equation containing unknown functions, their derivatives (differentials), and arguments is called a **differential**. In the general case of the function u (x, y, z, t...) the differential equation is

$$F\left(u, x, y, z, \dots, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}, \dots, \frac{\partial^i u}{\partial x^i}, \dots\right) = 0$$

With variables and their derivatives there can be both variable and constant coefficients. The order of the higher derivatives included in a differential equation determines the order of the equation. A distinction is made between equations of the first, second and higher orders. An equation of a function of one variable is called an **ordinary equation**, and an equation of functions of many variables is called a **partial derivative equation**. To solve a differential equation means to find its **integral**, which is one or more values of the unknown function that turns the equation into an identity. The operation of finding the integral of an equation is called its **integration**. Since integration is ambiguous and its result is defined with an accuracy up to arbitrary constants of integration,

the integral always contains these constants, and their number is equal to the order of the equation. To find an unambiguous solution of these problems, it is necessary to define in addition the equations defining the conditions of behavior of the function or its derivatives on the boundary of the domain of the desired solution (**boundary conditions**) and at the initial moment of time (**initial conditions**). Such problems are called **boundary problems** or **Cauchy problems**.

Integrating a differential equation is always a challenging creative task. However, there are a number of standard methods that make this task easier.

2.5.1. Methods for integrating first-order equations

Variable separation method. If the equation can be reduced to the form $M(x)\,K(y) = P(x)\,Q(y)$ then its variables can be divided and the left and right parts integrated separately. Suppose, for example, we are given the equation

$$xy' + y = 0.$$

It can be written in the form $xdy = -ydx$ and divided by its variables

$$\frac{dy}{y} = -\frac{dx}{x}.$$

By integrating, on the left and on the right, we obtain

$$\ln y = -\ln x + \ln C,$$

from which

$$y = \frac{C}{x} \tag{2.16}$$

In some cases, the equation can only be separated after appropriate transformations. An example is the homogeneous equation.

Let the following equation be given

$$M(x, y) + N(x, y)\, y' = 0$$

provided that the functions M and N are homogeneous with respect to their variables of the same degree. Homogeneity, for example of M, means that

$$M(ax, ay, \ldots) = a^n M(x, y, \ldots).$$

In this case the equation can be reduced to the form with separating variables by substituting $u = \frac{y}{x}$. For example, the equation

$$y^2 + x(x - y)\, y' = 0.$$

Here $M(x, y) = y^2$, and $N(x, y) = x(x - y)$ are homogeneous by definition functions (for example, for $N = ax(ax - ay) = a^2 x(x - y)$). We apply the substitution $y = ux$, then

$$\frac{dx}{x} + (1 - u)\frac{du}{u} = 0$$

Integration of this equation gives

$$\ln x + \ln u - u = \ln C.$$

By inverse substitution of the value of u we obtain

$$y = Ce^{\frac{y}{x}} \tag{2.17}$$

Total differential method. If the left side of the equation

$$Mdx + Ndy = 0 \tag{2.18}$$

can be represented by the total differential of some function u (see 2.14), the equation takes the form

$$du = 0,$$

and its integral

$$u(x, y) = C$$

It follows that

$$u = \int_{x_0}^{x} M(\varphi, y)d\varphi + \int_{y_0}^{y} N(x_0, \tau)d\tau \quad (2.19)$$

If the left part of equation 2.18 is not a total differential, then an integrating factor μ (x, y) is introduced, by multiplying by which (2.18) turns into an equation in total differentials. Substituting this factor into equation 2.14 leads to the equation for the factor μ

$$N\frac{\partial \ln \mu}{\partial x} - M\frac{\partial \ln \mu}{\partial y} = \frac{\partial M}{\partial y} - \frac{\partial N}{\partial x} \quad (2.20)$$

Let's consider a simple example. Solve the equation.

$$(x^2 + y)dx - x\,dy = 0. \quad (2.21)$$

In the given equation $M = x^2 + y$, and $N = -x$. Accordingly

$$\frac{\partial M}{\partial y} = 1, \frac{\partial N}{\partial x} = -1.$$

In other words, the total differential condition is not satisfied. Since the integrating factor from equation 2.20 cannot be found in the general case, we simplify equation 2.20, assuming that it is only a function of x, then according to (2.20) we obtain the equation

$$d \ln \mu = -2\frac{dx}{x} \quad (2.22)$$

Solving the resulting equation gives

$$\mu = \frac{1}{x^2}.$$

Substituting the resulting value of the integrating factor into M and N allows us to obtain the integral of the original equation from Equation 2.19 in the following form

$$x - \frac{y}{x} = C.$$

Integration of linear equations. **Linear equations** are differential equations in which the unknown function and its derivative are in the first

degree. The linear equation in general case is written in the following form

$$y' + P(x)y = Q(x) \tag{2.23}$$

It is easy to show that a linear equation can be reduced to an equation in total differentials by an integrating factor $\mu = e^{\int P dx}$. The general integral of the linear equation is

$$y = e^{-\int P dx}\left[\int Q e^{\int P dx} + C\right] \tag{2.24}$$

Other methods. Nonlinear equations of the first order can be solved only in partial cases. For example, if one particular solution is known or guessed (i.e. solution y_1 for some value of the integration constant C), then the general solution of the so-called Riccati equation of the form

$$y' = P(x)y^2 + Q(x)y + R(x)$$

is found by replacing variables $y = y_1 + \frac{1}{z}$, which reduces the original equation to a linear equation. The equation $F = 0$, not solved with respect to the derivative, can be solved if at some point (x_0, y_0) the equation $F(x_0, y_0, p) = 0$, where $p = y'$, has n valid roots p_i and can be solved with respect to the derivative. Reference books on higher mathematics and, in particular, on differential equations, provide a number of other cases of integration of first-order equations.

2.5.2. Methods of integration of higher order equations

The lowering of order. In some cases it is possible to lower the order of the equation, including when

- equations do not explicitly contain the argument x - by substituting variables $y' = p$; $y'' = p\frac{dp}{dy}$ etc.

- equations do not explicitly contain the function y - by substituting variables $y' = p$;

- equations are homogeneous with respect to the function and its derivatives - by replacing the variables $y' = yz$.

Linear equations. **Linear equations** of higher orders with constant coefficients a_i are equations of the following form

$$y^{(n)} + a_1 y^{(n-1)} + a_2 y^{(n-2)} + \cdots = F(x) \quad (2.25)$$

An equation with a nonzero right-hand side is called **inhomogeneous**. The equation with the right zero part is **homogeneous**. The solution to a inhomogeneous equation is the sum of the solution to the homogeneous one plus the partial solution to the non-homogeneous one. The solution of the homogeneous equation is found according to the relation

$$y = \sum C_i e^{r_i x} \quad (2.26)$$

where C_i are integration constants. Their number is equal to the order of the equation

r_i are roots of the characteristic equation, which are obtained from the algebraic equation by replacing the derivatives in the differential equation with appropriate degrees.

In the case of multiple roots of k-th multiplicity, solutions of the homogeneous equation are also linear combinations of

$$C_{i+k-1} x^{k-1}$$

For complex roots the Euler formula is used, according to which

$$e^{jrx} = \cos rx + j \sin rx.$$

The solution of an inhomogeneous equation is performed by the method of variation of constants given in the reference books. In the case of a special, for example, right transcendental part, a partial solution of an non-homogeneous equation is sought in the form of the right part. Let's consider an example of solving the most common in practice linear equations of the second order with constant coefficients and the special right part

$$\frac{d^2y}{dx^2} - 6\frac{dy}{dx} + 8y = e^{2x} \tag{2.27}$$

The characteristic equation

$$\begin{gathered} r^2 - 6r + 8 = 0 \\ r_1 = 4;\ r_2 = 2. \end{gathered} \tag{2.28}$$

Solving a homogeneous equation

$$y_0 = C_1e^{4x} + C_2e^{2x}.$$

We are looking for a partial solution of the inhomogeneous equation in the form of the right-hand side

$$y_1 = Axe^{2x}$$

$$\frac{dy_1}{dx} = 2Axe^{2x} + Ae^{2x};\ \frac{d^2y}{dx^2} = 4Axe^{2x} + 4Ae^{2x}.$$

Substitution gives the expression

$$4\,Axe^{2x} + 4\,Ae^{2x} - 12\,Axe^{2x} - 6\,Ae^{2x} + 8\,Axe^{2x} = e^{2x},$$

from which

$$A = -\frac{1}{2}$$

Thus, the solution to the equation has the following form

$$y = C_1e^{4x} + \left(C_2 - \frac{x}{2}\right)e^{2x} \tag{2.29}$$

Other methods for integrating ordinary differential equations can be found in special reference books on higher mathematics.

2.5.3. Equations in partial derivatives

Equations of the first order. As in the case of ordinary higher-order equations, we distinguish between linear and nonlinear, homogeneous and inhomogeneous partial differential equations. Let's consider solving linear equations of the following form

$$X_1 \frac{\partial z}{\partial x_1} + X_2 \frac{\partial z}{\partial x_2} + \cdots = Y. \quad (2.30)$$

The coefficients X_i, like the function z, are functions of n variables. The problem of integration of a linear equation is equivalent to the problem of integration of a system of first order ordinary equations

$$\frac{dx_1}{X_1} = \frac{dx_2}{X_2} = \cdots = \frac{dz}{Y}.$$

The problem of integrating a nonlinear equation is more complicated and will not be considered, since it is rarely applied in practice.

Problems in Mathematical Physics. Wave Equation. In physics, we most often have to solve the Cauchy problem for second-order partial differential linear equations with or without the right-hand side. These problems are called mathematical physics problems, and methods of their solution are studied in a special mathematical discipline called "Methods of Mathematical Physics". In mathematics, as we know, it is very common to use symbolic notations for the set of mathematical operations that must be performed on the arguments in order to obtain a function. These symbols are called **operators**. Examples of simple operators are signs of mathematical operations +, –, $\frac{d}{dx}$, $\sum$, $\int$, etc., symbols of trigonometric functions sin, cos, tan, etc., logarithmic symbols log, lg, ln, etc. Differential operators of the first, second and higher orders are also widespread. The following operators are the most used in the methods of mathematical physics:

nabla, or Del operator is a symbolic vector with following components

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right), \quad (2.31)$$

Laplace operator

$$\Delta = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right),$$

d'Alembert operator

$$\square = \frac{\partial^2}{\partial t^2} - a^2\Delta,$$

operators grad, div, rot, and so on.

Let us consider some simple problems of mathematical physics.

Wave Equation. The wave equation is a linear equation of vibration propagation in a homogeneous continuous medium. In the general case the wave equation has the form

$$\square = \frac{\partial^2 u}{\partial t^2} - a^2\Delta \mathrm{u} = \mathrm{Q(x, y, z, t)} \tag{2.32}$$

The function *u* (*x*, *y*, *z*, *t*), for which the wave equation is written is called the **wave function**. The wave equation describes almost all kinds of small vibrations in a solid material medium - longitudinal vibrations in gases, liquids, solids, transverse vibrations in strings, on the surface of water, and so on. The Cauchy problem for the case of a homogeneous equation is given by initial and boundary conditions for the function and its derivative. Particular cases of wave equations are: the stationary equation that takes place for the wave function independent of time (an example is the Schrödinger stationary equation in quantum mechanics), homogeneous one-dimensional equations (string equation), nonlinear equations, etc.

Let us consider, for example, the problem of free vibrations of a string. This is, firstly, a one-dimensional problem, since the wave arising in the string propagates only along the string direction, in the space of one dimension, coinciding, for example, with the Ox axis. Since the string, according to the condition, performs free vibrations, the problem is described by the homogeneous wave equation at the corresponding initial and boundary conditions, namely

$$\left.\begin{array}{l} \dfrac{\partial^2 u}{\partial t^2} - a^2 \dfrac{\partial^2 u}{\partial t^2} = 0 \\ u|_{t=0} = f(x), \qquad \left.\dfrac{\partial u}{\partial x}\right|_{t=0} = \varphi(x), \qquad u|_{x=0}^{x=1} = 0 \end{array}\right\} \tag{2.33}$$

Linear partial differential equations of higher orders, especially in the case of homogeneous equations, allow, as a rule, the substitution of

$$u(x_1, \ldots, x_n, t) = f_1(x_1) \cdot \ldots \cdot f_n(x_n) \cdot f_t(x_t) \quad (2.34)$$

Applying the above substitution in the form $u = X(x)\ T(t)$ for Equation 2.33 under the specified initial and boundary conditions and dividing the variables, we obtain a system of two ordinary homogeneous linear equations of second order

$$\frac{d^2X}{dx^2} + \lambda^2 X = 0;\ \frac{d^2T}{dt^2} + \lambda^2 a^2 T = 0,$$

from which

$$\left.\begin{aligned} X &= C_1 e^{j\lambda x} + C_2 e^{-j\lambda x} \\ T &= C_3 e^{j\lambda at} + C_4 e^{-j\lambda at} \end{aligned}\right\}$$

It follows from the boundary conditions that

$$X = C \sin \lambda x,\ \sin \lambda l = 0.$$

From the second equation it follows that $\lambda_n = \frac{n\pi}{l}$.

Substituting these values using the Euler formula, we find that

$$u_n = \left[a_n \cos\frac{\pi a n}{l} t + b_n \sin\frac{\pi a n}{l} t\right] \sin\frac{\pi n}{l} x \quad (2.35)$$

General solution of the equation is

$$u = \sum_{n}^{\infty} u_n \quad (2.36)$$

Satisfying the initial conditions of the problem, we find a_n and b_n as coefficients of the Fourier series, replacing for a_n, φ by f and T by l (see 2.7).

The problem of consideration of longitudinal vibrations of a rod differs from the considered problem of transverse vibrations of a string only by boundary conditions, according to which one end of the rod is free and the other end performs longitudinal vibrations so that the rate of change of its wave function along the length with the change in deviation x is proportional to the acting force F

$$\left.\frac{\partial u}{\partial x}\right|_{x=1} = kF.$$

The following tasks are more complicated ones:

- heat propagation in a homogeneous medium, the equation

$$\frac{\partial u}{\partial x} - a^2 \Delta u = Q(x,t),$$

- Laplace and Poisson equations, which are given by the potential theory equations, respectively; the homogeneous equation ($\Delta u = 0$ - Laplace equation) and the inhomogeneous equation ($\Delta u = -4\pi\rho$ - Poisson equation; here ρ is a point function, such as charge density);

- the propagation of electric current over wires (the problem of telegraphers), etc.

The solutions to these problems are given in reference guides on methods of mathematical physics.

Appendix 3. Fundamentals of Field Theory

A region of space, each point of which is associated with a value of some quantity, is called a **field** of that quantity. In the general case we should talk about **tensor fields**. The numerical axis is, for example, a field of real numbers, and the coordinate plane is a field of complex numbers. However, most often the concept of a field is used to describe physical quantities. These are, for example, fields of stress and shear rate tensors in hydroaerodynamics, fields of temperatures and pressures in thermodynamics, fields of velocities and forces in mechanics, fields of electric potential, electric, magnetic and electromagnetic fields in electrodynamics, and so on. Particular cases of tensor fields are scalar and vector fields.

A distinction is made between flat, plane-parallel and spatial fields. A **flat field** is a field defined only for points of some plane. A field in a plane perpendicular to some chosen direction, such as the normal, is

called **plane-parallel field**. The field can also be called a function of the point. For example, a scalar field U can be written as a scalar function of a vector argument $U(\vec{r})$.

Spatial fields are defined as functions of three coordinates in Cartesian, cylindrical or spherical coordinate systems. Flat and plane-parallel fields are defined as functions of two variables in Cartesian and polar coordinate systems. Linear fields are defined as a function of one variable in a linear coordinate system.

A scalar field can be centrally symmetric (central) or axisymmetric (axial). A **centrally symmetric** or spherically symmetric field is a scalar field whose values are equal for all points of space that are equidistant from some point called the **center**. The value of the field *U* at the point *M* in this case depends only on the distance from this point to the center. If the field values are the same for all points equidistant from some straight line (field axis), then the field is called **cylindrical-symmetric**, axial or cylindrical.

It is assumed that the field can have a source. So, for example, the source of the gravitational field is mass, the source of the electrostatic field is an electric charge, the source of the magnetic field is an electric charge moving with a constant speed, which is called a current, the source of the electromagnetic field is an electric charge moving with acceleration, etc. In the most general case, sources are called charges. Charges are characterized by their magnitude. A distinction is made between extended and point charges. A point charge is a charge whose size (extent) can be neglected compared to the distance to the field point in question. It is assumed that the charges of the same-name field interact with each other with some force. This force is called the field force. Known field forces are, for example, the gravitational force of gravitation, the Coulomb force of the interaction of electric charges, forces that cause the transfer of heat, material particles, etc. The force of a field, referred to the value of its charge, is called the **field strength**. The field strength is a vector quantity. It represents the force characteristic of the field at each of its points. Electric and magnetic field strengths are

widely known. The gravity field strength is the acceleration vector of gravity, etc. The field, which is characterized by strength, is called a **force field**.

Consider some force field with strength $\vec{E}$ in the Cartesian coordinate system. Let's call the **field lines** the non-intersecting lines drawn through the field points so that their direction at each point coincides with the direction of the field strength vector at that point. The density of field lines in the vicinity of a given point must be proportional to the magnitude of the field strength vector. According to this definition, field lines can be seen as a geometric image or picture of the field.

Let us choose an arbitrary point *A* (*x*, *y*, *z*) in the field. Let us set the scalar function *φ* (*x*, *y*, *z*) at this point so that

$$\delta\varphi = -\left(\vec{E}, d\vec{s}\right), \tag{3.1}$$

where *ds* (*dx*, *dy*, *dz*) is an arbitrary element of the line drawn through point A.

The sign *δ* means that the function element *φ* in the general case is not a complete differential of the field point coordinates. As a special case, a field is called **conservative** if at each of its points there exists a scalar function *φ* whose differential is total. Otherwise, the field is called **nonconservative** or **vortex field**.

For a conservative field, according to its definition,

$$\delta\varphi = d\varphi = \frac{\partial\varphi}{\partial x}dx + \frac{\partial\varphi}{\partial y}dy + \frac{\partial\varphi}{\partial z}dz. \tag{3.2}$$

On the other hand, according to the definition of the scalar product of two vectors and equation (3.2)

$$E_x = -\frac{\partial\varphi}{\partial x};\ E_y = -\frac{\partial\varphi}{\partial y};\ E_z = \frac{\partial\varphi}{\partial z}. \tag{3.3}$$

The vector $\vec{a}$, whose components are equal respectively $\frac{\partial\varphi}{\partial x}; \frac{\partial\varphi}{\partial y}; \frac{\partial\varphi}{\partial z}$ to some scalar function φ is called the gradient of this function or the gradient of the scalar field ($\vec{a} = \nabla\varphi$). It follows that for a

conservative field there exists a scalar function φ such that the strenght vector of this field

$$\vec{E} = \text{grad}\,\varphi \tag{3.4}$$

It follows from the definition of the function φ that from the physical point of view it has the meaning of work. If a charge of the same nature as the source is placed at a given point in the force field, the work of the charge movement is done under the action of the field forces. Such a charge is called a test charge. Thus, the differential of the above function expresses the work of the field strength vector to move a unit of sample charge along the elementary path *ds* in the direction of the vector $d\vec{s}$. The minus sign means that the work is performed against the field forces. In the case of a conservative field, this function is called a **potential**.

Let us denote the work of moving a unit of a test charge in a conservative field by φ, then, as we know, the work *A* of moving a charge q from point 1 to point 2 of the field along the path *ds* under the action of the strength $\vec{E}$ is

$$A = q\int_1^2 (\vec{E}, d\vec{s}). \tag{3.5}$$

For the conservative field

$$A = q\int_1^2 d\varphi = q(\varphi_1 - \varphi_2). \tag{3.6}$$

In other words, the work of moving a test charge under the action of conservative field forces does not depend on the shape of the path, but is determined by the potentials at the starting and ending points of motion. This means that the scalar potential of a conservative field can be defined as the energy characteristic of the field at each of its points.

In the general case, the values of some field characteristic may be the same for some set of field points. Such a set of points forms a surface in space, which is called a **level surface**. Surfaces of the conservative field level for which φ = const are called **equipotential**. In the case of a

flat field, the level surfaces will be reduced to **level lines**. When the field strength vector $\vec{E}$ is perpendicular to the displacement $d\vec{s}$, the field is called central. The differential of the central potential field is zero, and its equipotential surfaces are concentric spheres or, for a flat field, circles. The field lines in this case are radial.

Let us consider the mathematical meaning of the gradient vector.

3.1. Gradient of the scalar field

Let us consider a scalar field $\varphi(\vec{r}) = \varphi(x, y, z)$, for example, the field of electrostatic potential. Let us call the **derivative along the direction** *s* of a function φ at point M_0 (x_0, y_0, z_0) a quantity $\frac{\partial \varphi}{\partial s}$ equal to the limit of the ratio of the increment of the function φ to the displacement Δs along the direction *s* in the neighborhood of point M_0, when it shrinks to that point. In other words, the derivative along the direction

$$\left(\frac{\partial \varphi}{\partial s}\right)_0 = \lim_{\Delta s \to 0} \frac{\Delta \varphi}{\Delta s} \tag{3.8}$$

It follows from this definition that the derivative along the direction characterizes the rate of change of a function in a given direction and at a given point.

Let us draw a normal $\vec{n}$ and an arbitrary direction $\vec{s}$ to point M_0 of the level surface φ_0. Let us consider the points of intersection of M_n and M_S of the normal *n* and direction *s* with the level surface $\varphi_0 + \Delta\varphi$ (see Fig. 3.1).

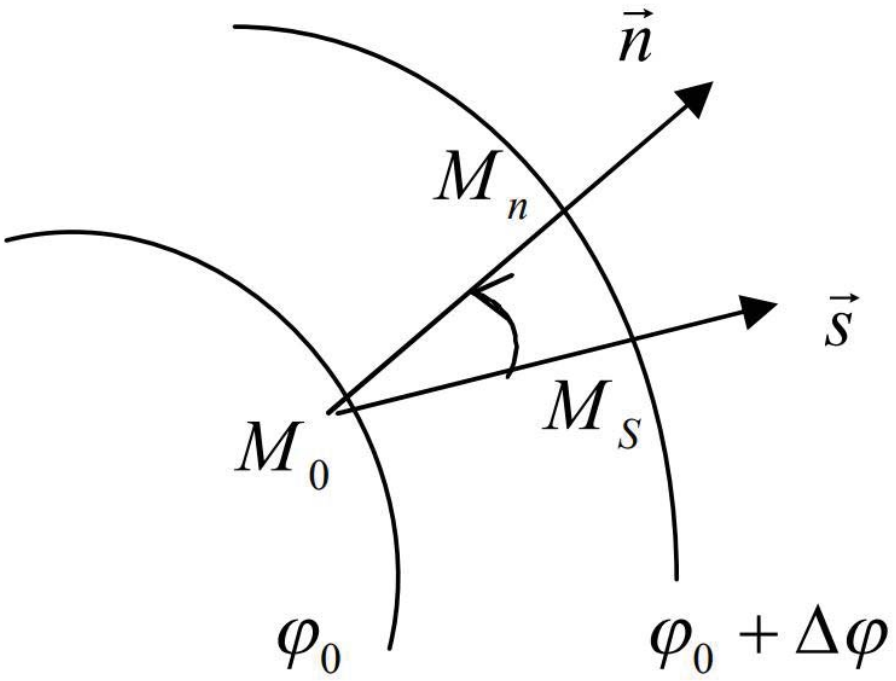


Figure 3.1.

According to the above definition

$$\left(\frac{\partial\varphi}{\partial s}\right)_0 = \lim_{M_0M_S\to 0} \frac{\varphi_s - \varphi_0}{M_0M_S} \tag{3.8}$$

Since the points M_s and M_n lie on the level surface, the values of the function in it φ_s and φ_n are equal to each other. On the other hand, as can be seen from Figure 3.1,

$$\begin{aligned} M_0M_S &= \frac{M_0M_n}{\cos(\vec{s},\vec{n})}, \\ \frac{\partial\varphi}{\partial s} &= \frac{\partial\varphi}{\partial n}\cos(\vec{s},\vec{n}). \end{aligned} \tag{3.9}$$

It follows that the rate of change of the function at a given point is maximal in the direction of the normal. Based on the above, let us introduce a special vector $\vec{b}$, directed along the normal to the level surface in the direction of increasing function, numerically equal to the derivative along the direction of the normal, and call it the **gradient** of the function. It is, in this way,

$$\vec{b} = \frac{\partial\varphi}{\partial n}\vec{n}_0 \tag{3.10}$$

$$\left|\vec{b}\right| = \frac{\partial \varphi}{\partial n} \tag{3.11}$$

Substituting 3.10 into 3.8 gives

$$\frac{\partial \varphi}{\partial s} = \left|\vec{b}\right| \cos(\vec{s}, \vec{n}). \tag{3.12}$$

It follows from 3.11 that the directional derivative is equal to the projection of the gradient vector onto this direction. Choosing coordinate axes of Cartesian coordinate system as directions, we can, therefore, write for components of gradient vector in Cartesian coordinate system

$$b_x = \frac{\partial \varphi}{\partial x};\ \ b_y = \frac{\partial \varphi}{\partial y};\ \ b_z = \frac{\partial \varphi}{\partial z}. \tag{3.13}$$

From (3.13) and (3.4) we see that

$$\vec{b} = \text{grad}\, \varphi \tag{3.14}$$

Components of the gradient (without proof): in the cylindrical coordinate system

$$\text{grad}\left(\frac{\partial}{\partial r}, \frac{1}{r}\frac{\partial}{\partial \theta}, \frac{\partial}{\partial z}\right);$$

and in the spherical coordinate system

$$\text{grad}\left(\frac{\partial}{\partial r}, \frac{1}{r}\sin\theta\,\frac{\partial}{\partial \theta}, \frac{1}{r}\,\frac{\partial}{\partial \psi}\right);$$

Applying the Nabla operator according to 2.31, we obtain

$$\text{grad}\, \varphi = \nabla \varphi.$$

It also follows from the definition of the gradient that the gradient of a function can be viewed as a certain operator that converts a **scalar to a vector**, or, more precisely, increases the rank of the tensor by one unit. Since the gradient is a differential operator, all rules of differentiation apply to it.

Let us calculate, for example, the gradient of the numerical value of the position vector $\vec{r}$ of some point *M* with respect to the point of reference O (see Fig. 3.2).

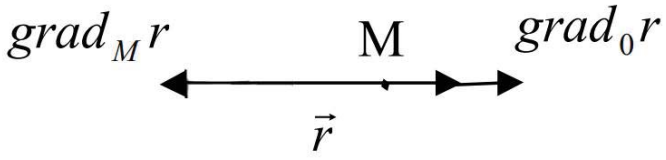


Figure 3.2.

When the position of point O is fixed, the maximum change in the length (numerical value) of the position vector r = OM occurs in the direction of motion of point M from O toward M, and $\text{grad}\, r = \frac{dr}{dr} = 1$, therefore, $\text{grad}_0\, r = \frac{\vec{r}}{r}$. In the case of a fixed position of the observation point and moving the reference point

$$\text{grad}_M\, r = \frac{\vec{r}}{r}$$

In the general case

$$\text{grad}\, f(r) = \frac{\partial f}{\partial r}\text{grad}\, r = \pm\frac{\partial f}{\partial r}$$

3.2. Flux of the Vector. Divergence.

Depending on the configuration of the field, its lines can be **closed** or **open**, including those going to infinity. Open lines begin at points called field **sources** and end at points called field **sinks**. The **flux** of vector field strength *dN* through the area *dS* is the value proportional to the area and to the value of the field strength vector perpendicular to the area. If a vector field is defined by some vector $\vec{a}\,(\vec{r})$, then according to this definition

$$dN = a_n dS \qquad (3.15)$$

where a_n is the projection of the vector $\vec{a}$ on the normal to the area at the given point.

According to the definition of the scalar product (see section 1.5.3 of Appendix 1)

$$dN = \left(\vec{a}, d\vec{S}\right) \tag{3.16}$$

Let us choose a point in the field localization space $\vec{a}$ and surround it with an elementary closed surface *dS*. Considering the specified point as the source of the field, let us determine the flux of the field strength vector through this surface. Let us conventionally assume that the flux entering the volume bounded by this surface is positive. The flux coming out of this volume is assumed to be negative. Let's choose a Cartesian coordinate system with the origin at a given point. In the space of the chosen reference system there are only 3 independent directions along the coordinate axes, and in each of these directions the surface is intersected by a flux of lines entering and leaving the volume. Thus, for example, the flux in the direction of the Ox axis is equal according to the above definition

$$dN_x = (a_{2x} - a_{1x})\, dydz.$$

Here the expression in parentheses is the increment of the function *a* (*x*, *y*, *z*) in the direction of the Ox axis and, according to the definition of the partial increment of the function (see Section 2.2), is equal to

$$\frac{\partial a_x}{\partial x} dx.$$

It follows that the total total flux through the closed area dS is

$$dN = \left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}\right) dxdydz \tag{3.17}$$

where *dx dy dz* = *dV* is the elementary volume covered by the area *dS*.

The scalar expression in parentheses of the relation (3.17) is called the **divergence** of the vector a and is denoted as follows div $\vec{a}$

$$\operatorname{div}\vec{a} = \left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}\right) \tag{3.18}$$

It follows from (3.18) that the divergence can also be written through the Nabla operator as follows

$$\operatorname{div}\vec{a} = (\nabla, \vec{a}) \tag{3.19}$$

In other words, divergence lowers the rank of the tensor by one unit.

Thus, finally

$$dN = \operatorname{div}\vec{a}\, dV. \tag{3.20}$$

From (3.17) and the definition of the derivative it also follows that

$$\operatorname{div}\vec{a} = \lim_{\Delta V \to 0} \frac{\int_{S'} a_n dS}{\Delta V} \tag{3.21}$$

From (3.21) we can also see that divergence can be viewed as a volume differentiation.

The flux through the finite surface *S*, covering the volume *V*, is equal to

$$N = \int_V \operatorname{div}\vec{a}\, dV. \tag{3.22}$$

Equation (3.22) shows that if the vector divergence is zero, the flux is zero, i.e. the field has neither sources nor sinks. It was shown above that for conservative fields $\vec{E} = -\operatorname{grad}\varphi$. It follows that the divergence of the conservative field strength vector is

$$\operatorname{div}\vec{E} = -(\nabla, \nabla\varphi) = -\Delta\varphi \neq 0.$$

This means that the conservative field lines are not closed, i.e., they start at the sources and end at the sinks. In the case when the field lines are closed, i.e. the field divergence is zero, the field, as mentioned above, is vortex. Hence the name of the introduced quantity is **divergence**.

Divergence is a differential operator, so it, like the gradient, obeys the rules of differentiation, although it also has some specific features. For example,

$$\text{grad div } \vec{a} = \nabla(\nabla, \vec{a}) \neq \nabla^2 \vec{a},$$

as it would be in a normal differentiation.

Comparing 3.16 and 3.20, we find that

$$\int_S a_n dS = \int_V \text{div } \vec{a}\, dV. \tag{3.23}$$

The relation (3.23) is known as the Gaussian formula.

In the cylindrical coordinate system

$$\text{div } \vec{a} = \frac{1}{r}\frac{\partial}{\partial r}(r a_x) + \frac{1}{r}\frac{\partial a_\theta}{\partial \theta} + \frac{\partial a_z}{\partial z}$$

In the spherical coordinate system

$$\text{div } \vec{a} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 a_r) + \frac{1}{r \sin\theta}\frac{\partial a_\varphi}{\partial \varphi} + \frac{1}{r \sin\theta}\frac{\partial}{\partial \theta}\sin\theta a_\theta.$$

3.3. Circulation of the Vector. Curl.

An important characteristic of a vector field is its circulation *C* (see Section 2.4). An example of circulation can be, in particular, the work of field forces to move a point (charge) along a closed loop.

It is easy to show that the circulation of the conservative vector field is zero. Indeed, as shown above, a conservative field $\vec{a}$ is described by a nonzero divergence and it is always possible to find a scalar function φ for which $\vec{a} = \text{grad } \varphi$. According to the definition of circulation

$$C = \oint_S a_S ds = \oint_S (\text{grad } \varphi, d\vec{s}) = \oint_S d\varphi = 0 \tag{3.24}$$

The result reflects the well-known fact that the work of the conservative field forces does not depend on the shape of the path, that is, it is determined only by the initial and final point of displacement. When moving along a closed loop, the positive work performed on the direct path is compensated by the negative work on the return path.

Vortex fields behave differently. The work of the vortex field forces does not depend on the direction of the loop traversal and is determined not only by the initial and end points of the displacement path, but also by its shape. Therefore, the work of the vortex field forces on a closed loop, and in the general case, the circulation of the vector of this field, is not equal to zero. To describe the vortex fields, a special vector related to circulation is introduced in connection with the above. This vector is called a **curl** and is denoted by rot. The rotor is introduced as follows. For simplicity, let the closed circulation loop of the vortex field vector $\vec{a}$ is a flat elementary rectangle ABCD with the center at the beginning of coordinates (see Fig. 3.3).

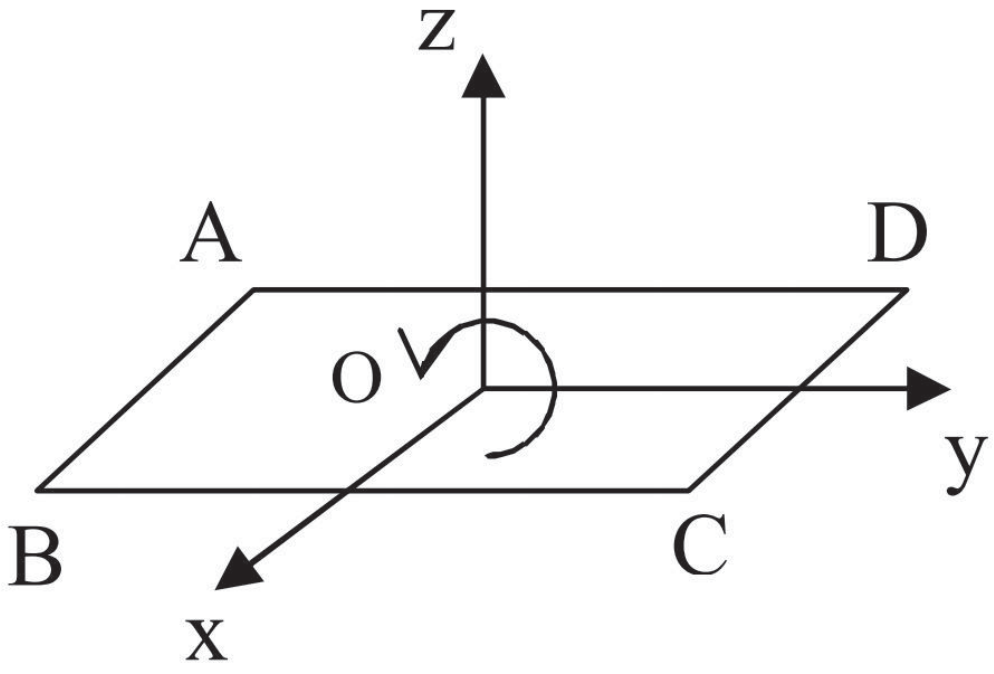


Figure 3.3.

Let us place this contour in the xOy plane so that its sides AB and CD, equal to dx, are parallel to the Ox axis, and its sides DA and BC, equal to du, are parallel to the Oy axis. The Oz axis in this case coincides

with its normal for a positive counterclockwise bypass of the contour. Circulation of the field vector along this contour

$$C = \oint_C a_S ds = \int_A^B a_{x_1} dx + \int_B^C a_{y_1} dy - \int_C^D a_{x_2} dx - \int_D^A a_{y_2} dy \tag{3.25}$$

According to the mean value theorem (see Section 2.2)

$$a_{x_1} - a_{x_2} = -\frac{\partial a_x}{\partial y} dy$$

and so on. Therefore, the circulation along the contour located in the xOy plane, the normal of which is the Oz axis, is described by the relation

$$C_z = \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right) dS \quad (dS = dxdy) \tag{3.26}$$

Similarly

$$\left.\begin{aligned} C_y &= \left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}\right) dS \quad (dS = dxdz) \\ C_x &= \left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}\right) dS \quad (dS = dydz) \end{aligned}\right\} \tag{3.27}$$

It follows from (3.26) and (3.27) that there is a vector $\mathrm{rot}\,\vec{a}$ with components C_x, C_y, and C_z that describes the circulation. This vector can be written using the following third-order determinant

$$\mathrm{rot}\,\vec{a} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a_x & a_y & a_z \end{vmatrix} = [\nabla, \vec{a}] \tag{3.28}$$

A comparison of 3.25 and 3.28 gives

$$\int_C (\vec{a}, d\vec{s}) = \int_S \left(\text{rot}\,\vec{a}, d\vec{S}\right) \tag{3.29}$$

The resulting ratio is called the **Stokes** formula or theorem. From the Stokes formula, in particular, another definition of the rotor follows, namely: "The curl of a vortex vector field is a vector whose normal component at a given point is equal to the derivative of the circulation of the field vector along the contour of an arbitrary area *dS*, taken by the size of this area". The surface passes through this point perpendicular to the curl.

Based on the obtained relations, it is easy to show that

$$\text{rot}\,\text{grad}\,\varphi = 0;\ \text{div}\,\text{rot}\,\vec{a} = 0;\ \text{rot}\,\text{rot}\,\vec{a} = \text{grad}\,\text{div}\,\vec{a} - \nabla^2\vec{a}. \tag{3.30}$$

It follows from the first relation (3.30) that the conservative field has no vortices. From the second relation it follows that the vortex field cannot have open lines. Finally, from the third equation it follows that the curl of the vortex field cannot be equal to zero.

Curl as a differential operator, obeys the rules of differentiation, although it has some peculiarities. In addition, unlike the gradient and divergence, it does not change the rank of the tensor and, in particular, converts a vector to a vector.

We leave it to the students to write down an expression for the curl in cylindrical and spherical coordinate systems.

The properties and relations for differential and integral operators of vector (tensor) analysis are described in more detail in special manuals on higher mathematics.

Appendix 4. International system of units of physical quantities SI

Table 4.1. Base units

The quantity		Unit	
		Name	Symbol
1	Length (linear dimension)	meter	m
2	Mass	kilogram	kg
3	Time	second	s
4	Electric current	amper	A
5	Thermodynamic temperature	kelvin	K
6	Luminous intensity	candela	cd
7	Amount of substance	mole	mol

Table 4.2. Derived units

The quantity		Unit		Expression (in SI based units)
		Name	Symbol	
1	Plane angle	radian	rad	$m \cdot m^{-1} = 1$
2	Solid angle	steradian	sr	$m^2 \cdot m^{-2} = 1$
3	Celsius temperature	degree Celsius	ºC	K
4	Frequency	herz	Hz	s^{-1}
5	Velocity, speed	-	-	$m \cdot s^{-1}$
6	Acceleration	-	-	$m \cdot s^{-2}$
7	Force	newton	N	$kg \cdot ms^{-2}$
8	Energy, work, heat	joule	J	$N \cdot m = kgm^2s^{-2}$
9	Momentum	-	-	$N \cdot s = kgms^{-1}$
10	Power	watt	W	Js-1
11	Pressure (stress)	pascal	Pa	$N \cdot m^{-2} = kgm^{-1}$
12	Luminous flux	lumen	lm	cd · sr
13	Illuminance	lux	lx	$lm \cdot m^{-2} = cd \cdot sr^{-2}$

14	Electric charge	coulomb	C	$A \cdot s$
15	Electric potential (voltage)	volt	V	$J \cdot C^{-1} = kgm^2s^{-3}A^{-1}$
16	Resistance	ohm	Ω	$V \cdot A^{-1} = kgm^2s^{-3}A^{-2}$
17	Capacitance	farad	F	$C \cdot V^{-1} = kg^{-1}m^{-2}s^4A^2$
18	Magnetic flux	weber	Wb	$V \cdot s = kgm^2s^{-2}A^{-1}$
19	Magnetic flux density	tesla	T	$Wb \cdot m^{-2} = kgs^{-2}A^{-1}$
20	Inductance	henry	H	$Ohm \cdot s = kgm^2s^{-2}A^{-2}$
21	Electrical conductance	siemens	S	$Ohm^{-1} = kg^{-1}m^{-2}s^3A^2$
22	Activity referred to a radionuclide	becquerel	Bq	s^{-1}
23	Absorbed dose of ionising radiation	gray	Gg	$J \cdot kg^{-1} = m^2s^{-2}$
24	Equivalent dose of ionising radiation	sivert	Sv	$J \cdot kg^{-1} = m^2s^{-2}$
25	Catalytic activity	katal	kat	$mole \cdot s^{-1}$

Table 4.3. Non-SI units

	Unit	Symbol	Value in SI units
1	minute	min	60 s
2	hour	h	60 min = 3600 s
3	day	d	24 h = 86400 s
4	month	-	30,4375 days = 2,6298 · 10^6s
5	year	-	365,25 days = 3.15576 · 10^7 s
6	light year	-	9,461 · 10^{15} m
7	parsec	p^s	3,086 · 10^{16} m
8	unified atomic mass unit (Dalton)	Da	1,6605402·10^{-27} kg
9	Electronvolt	eV	1,60217733·10^{-19} J
10	bel	B	-
11	neper	Np	-
12	litre	L	10^{-3} m^3
13	tonne	t	10^3 kg
14	astronomical unit	ua	1,49597870691·10^{11} m
15	nautical mile	-	1852 m
16	knot	-	mile · h^{-1}= =0,5144444 ms^{-1}
17	are	a	10^2 m^2
18	hectare	ha	10^4 m^2
19	bar	bar	10^5 Pa
20	angstrom	Å	10^{-10} m
21	barn	b	10^{-28} m^2
22	grad	°	1,74444·10^{-2} rad
23	minute	!	2,907·10^{-6} rad
24	second	"	4,845·10^{-8} rad
25	dioptry	dptr	m^{-1}
26	kW-hour	kW-h	3,6·10^6 J
27	V-amper	VA	1 W
28	amper-hour	A-h	3600 As

Notes

1. There is no dot placed after the unit designation.

2. The name of the unit in the text is written with a lowercase letter, and the designation of units named after scientists is written with a capital letter.

3. Fractional and multiple units are denoted with a prefix to the main designation.

4. Standards of units of physical quantities are used to ensure the uniformity of measurements. Standards are stored in special conditions and are subject to periodic testing. A distinction is made between primary and secondary, state and international standards. Basic units of measurement are reproduced, as a rule, with the help of primary standards in accordance with their definition. Secondary standards reproduce derived units. The definition of measurement standards is given in reference books on metrology.

Appendix 5. Physical constants

5.1. Fundamental physical constants

1	Gravitational constant	$G = 6{,}67 \cdot 10^{-11}\ Nm^2kg^{-2}$
2	Speed of light in a vacuum	$c = 2{,}997... \cdot 10^8 \approx 3 \cdot 10^8\ ms^{-1}$
3	The Planck's constant	$\hbar = h/2n\pi \approx 1{,}054 \cdot 10^{-34}\ J \cdot s$
4	Electron rest mass	$m_0 \approx 9{,}109 \cdot 10^{-31}\ kg$
5	Proton rest mass	$m_p \approx 1{,}673 \cdot 10^{-27}\ kg$
6	Neutron rest mass	$Mp \approx 1{,}675 \cdot 10^{-27}\ kg$
7	Electron charge	$e = 1{,}602 \cdot 10^{-19}\ C$
8	Avogadro's constant	$N_A \approx 6{,}022 \cdot 10^{23}$
9	Boltzmann's constant	$K \approx 1{,}381 \cdot 10^{-23}$
10	Vacuum permittivity	$\varepsilon_0 \approx 8{,}854 \cdot 10^{-12}\ Fm^{-1}$
11	Vacuum magnetic permeability	$\mu_0 \approx 12{,}57 \cdot 10^{-8}\ Hnm^{-1}$
12	Gas constant	$R \approx 8{,}314\ J \cdot mole$

5.2. Astronomical constants

1	Distance of the Earth from the Sun (average) (astronomical unit)	$\sim 1{,}496 \cdot 10^{11}$ m
2	Mass of the Sun	$\sim 1{,}989 \cdot 10^{30}$ kg
3	Radius of the Sun	$\sim 6{,}9599 \cdot 10^8$ m
4	Mass of the Earth	$\sim 5{,}976 \cdot 10^{24}$ kg
5	Radius of the Earth	$\sim 6{,}378 \cdot 10^6$ m
6	Mass of the Moon	$\sim 7{,}35 \cdot 10^{22}$ kg

L. Preigerman, M. Brooke, V. Mikhalev-Volookov

MODERN PHYSICS COURSE

New Approaches
to Explaining the Physical
Picture of the World

Volume 2

Signed to the press on 03.02.2023. Format 60x84 1/16.
Offset printing. Offset paper.
Usl. P. L. 31.75. Edition of 10 copies .

Printed in the printing house of LLC "New Project"
Bryansk, Uritsky lane, 18
Tel.: (4832) 74-47-86